GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 2719.24 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-3
Perfect score: 2170
Sequence: 1 ctttccctcgctgggatcgc.....aaaaaaaaaaaaaaaaaaaa 2170

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2139	98.6	2139	10	AB015426	AB015426 Mus muscu
2	1588.2	73.2	2156	10	AB049819	AB049819 Rattus no
3	1309.2	60.3	1705	10	AF230460	AF230460 Cricetulu
4	1206.8	55.6	3019	9	AB023021	AB023021 Homo sapi
5	1205	55.5	2501	9	HSA238701	AJ238701 Homo sapi
6	1162.4	53.6	80247	9	AL512406	AL512406 Human DNA
7	1049.6	48.4	1128	10	AF345993	AF345993 Rattus no
c 8	883	40.7	71396	2	AC100674	AC100674 Mus muscu
9	740	34.1	1707	5	AB035906	AB035906 Gallus ga
.. 10	740	34.1	71396	2	AC100674	AC100674 Mus muscu
11	675.2	31.1	1080	5	AB035905	AB035905 Xenopus l
12	269	12.4	1567	5	AB023627	AB023627 Danio rer
13	230.4	10.6	1832	5	AB023628	AB023628 Danio rer
14	220.6	10.2	1264	5	AF288369	AF288369 Gallus ga
15	206.6	9.5	1092	5	AB035907	AB035907 Xenopus l
16	175.2	8.1	2657	5	GGU73678	U73678 Gallus gall
17	157	7.2	1429	10	CGU78737	U78737 Cricetulus
18	157	7.2	1462	10	AF090450	AF090450 Cricetulu
19	155.4	7.2	1718	10	AF090449	AF090449 Cricetulu
20	135.6	6.2	1044	5	AB035908	AB035908 Gallus ga
21	134.8	6.2	2293	4	BTA134FTG	X87810 B.taurus al
22	134.8	6.2	12607	4	BTA132772	AJ132772 Bos tauru
23	134.2	6.2	1158	10	AB039126	AB039126 Mus muscu
24	133.4	6.1	1113	9	AF345884	AF345884 Saimiri s
25	133	6.1	1119	9	AF345881	AF345881 Macaca mu
26	132.6	6.1	1170	10	AB039127	AB039127 Mus muscu
27	132.6	6.1	1170	10	AB039131	AB039131 Mus muscu
28	132.6	6.1	1615	10	MMU33457	U33457 Mus musculu
29	131.4	6.1	1170	10	AB039133	AB039133 Mus spici
30	131	6.0	1170	10	AB039124	AB039124 Mus muscu
31	131	6.0	1170	10	AB039125	AB039125 Mus muscu
32	131	6.0	1170	10	AB039128	AB039128 Mus muscu
33	131	6.0	1170	10	AB039129	AB039129 Mus muscu
34	131	6.0	1170	10	AB039130	AB039130 Mus muscu
35	131	6.0	1170	10	AB039132	AB039132 Mus muscu
36	130.8	6.0	1302	10	AB049938	AB049938 Rattus no
37	130.8	6.0	169446	2	AC097778	AC097778 Rattus no
38	130	6.0	1613	4	BTA132776	AJ132776 Bos tauru
39	130	6.0	1619	4	BTA132775	AJ132775 Bos tauru
40	130	6.0	1644	4	BTA132774	AJ132774 Bos tauru
41	130	6.0	1873	4	BTA132773	AJ132773 Bos tauru
42	129.8	6.0	1162	9	PTY14033	Y14033 Pan troglod
c 43	129.8	6.0	80893	9	AC024592	AC024592 Homo sapi
44	129.8	6.0	187488	2	AC010454	AC010454 Homo sapi
c 45	128.8	5.9	169893	2	AC021165	AC021165 Homo sapi

ALIGNMENTS

RESULT 1
AB015426
LOCUS AB015426 2139 bp mRNA linear ROD 23-JUN-1999
DEFINITION Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete cds.
ACCESSION AB015426
VERSION AB015426.1 GI:3702718
KEYWORDS Fut9; alpha1,3-fucosyltransferase IX.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K. and Narimatsu,H.
TITLE Expression cloning and characterization of a novel murine alpha1, 3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15) epitope in brain and kidney
JOURNAL J. Biol. Chem. 273 (41), 26729-26738 (1998)
MEDLINE 98434588
REFERENCE 2 (bases 1 to 2139)
AUTHORS Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN
```

Query Match 98.6%; Score 2139; DB 10; Length 2139;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
Db       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60

Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
Db      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120

Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
Db     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180

Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
Db     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240

Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
Db     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300

Qy     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
Db     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360

Qy     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420
Db     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420

Qy     421 TCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAG 480
Db     421 TCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAG 480

Qy     481 CAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACC 540
Db     481 CAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACC 540

Qy     541 CCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCA 600
Db     541 CCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCA 600

Qy     601 GATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTG 660
Db     601 GATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTG 660

Qy     661 CCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGG 720
Db     661 CCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGG 720

Qy     721 GTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTC 780
Db     721 GTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTC 780

Qy     781 GGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTT 840
Db     781 GGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTT 840

Qy     841 TCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTG 900
Db     841 TCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTG 900

Qy     901 GCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCA 960
Db     901 GCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCA 960

Qy     961 GCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 1020
Db     961 GCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 1020

Qy    1021 AAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTT 1080
Db    1021 AAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTT 1080

Qy    1081 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1140
Db    1081 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1140

Qy    1141 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTC 1200
Db    1141 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTC 1200

Qy    1201 CATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA 1260
Db    1201 CATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA 1260

Qy    1261 AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATAT 1320
Db    1261 AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATAT 1320

Qy    1321 TTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCC 1380
Db    1321 TTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCC 1380

Qy    1381 TGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAACAT 1440
Db    1381 TGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAACAT 1440

Qy    1441 TGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTACTTGTTTCTACATCA 1500
Db    1441 TGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTACTTGTTTCTACATCA 1500

Qy    1501 ATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTT 1560
Db    1501 ATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTT 1560

Qy    1561 TGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTG 1620
Db    1561 TGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTG 1620

Qy    1621 AATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGATAGAG 1680
Db    1621 AATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGATAGAG 1680

Qy    1681 TTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATAT 1740
```

```
Db    1681 TTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATAT 1740

Qy    1741 GAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 GAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGA 1800

Qy    1801 TTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTT 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 TTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTT 1860

Qy    1861 AGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 AGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTT 1920

Qy    1921 TGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 TGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCT 1980

Qy    1981 TATAACAATAAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATACCATTTAT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 TATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATACCATTTAT 2040

Qy    2041 AGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCATCCATATTTAAGTTT 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 AGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCATCCATATTTAAGTTT 2100

Qy    2101 TATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTG 2139
           |||||||||||||||||||||||||||||||||||||||
Db    2101 TATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTG 2139
```

```
RESULT 2
AB049819
LOCUS       AB049819                2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN
```

```
  Query Match             73.2%;  Score 1588.2;  DB 10;  Length 2156;
  Best Local Similarity   90.8%;  Pred. No. 1.8e-259;
  Matches 1748;  Conservative    0;  Mismatches  168;  Indels    9;  Gaps    5;

Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
           |||||||| |||||||||| ||||||||||||||| ||| | ||||||   ||| |||||
Db     234 CTTTCCCTTGCTGGGATCGATTTAGAATGTGATAAGTCAAGTATCTGAA--CACAGTGGA 291

Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
           ||||||| |||||||||| |||||||||| |||||||||||  ||||||||||||||
Db     292 GTAGTTCAACACTCATCTCAGCTTTCCATGATATGTTCTCTACCGTGAAAAATTATGACC 351

Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     352 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGGGCTGCTTC 411

Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
           |||||||||| || || || ||||||||||||||||||||||||||||||||||||||||
Db     412 GTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 471

Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
           |||||||| |||||||||||||||||||||||||||||  ||||||||||||||||||
Db     472 TCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATTATTTTAAT 531

Qy     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
           |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     532 GAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 591

Qy     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420
           ||||| |||||||| |||||||||||||||||||||| |||||||| ||||||||||||
Db     592 CAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGTACAACAAA 651

Qy     421 TCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAG 480
           |||||||||||||||||| ||||| |||||||||||||||||||||||||||||||||||
Db     652 TCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTACCTCAG 711

Qy     481 CAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACC 540
           |||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||
Db     712 CAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCACTCACACC 771

Qy     541 CCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCA 600
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     772 CCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCA 831

Qy     601 GATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTG 660
           ||||| |||||||||||| ||||||||||||||||||||  ||||||||||||||||||
Db     832 GATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTTGAAGTG 891

Qy     661 CCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGG 720
           |||||||| |||||||| |||||||||||||||||||||||||| || ||||||||||||
Db     892 CCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCATGCCAGG 951

Qy     721 GTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     952 GTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGCAAGCATTT 1011

Qy     781 GGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTT 840
           |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1012 GGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTT 1071

Qy     841 TCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTG 900
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1072 TCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTG 1131
```

```
Qy    901 GCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCA 960
          |||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1132 GCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCA 1191

Qy    961 GCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 1020
          |||||||||||||||||||||||||||||||| |||||||||| |||||||||| |||||||
Db   1192 GCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGTGAGTTGGCAAAATATCTG 1251

Qy   1021 AAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTT 1080
          ||||| |||||||||||||||||||| || |||||||||| |||||||| |||||||||||
Db   1252 AAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGAAGGATTTT 1311

Qy   1081 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1140
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1312 ACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATCATGTAAAA 1371

Qy   1141 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTC 1200
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   1372 AGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAAAGTGTC 1431

Qy   1201 CATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA 1260
          |||||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   1432 CATCATCGTCACAGTGAAATAGATTAATGAGACACCATTCAAGTGTTCAGGATAAGAAGA 1491

Qy   1261 AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATAT 1320
          |||||||||||||||||| ||| ||||| |||||||| |||||||||||||||||||||||||
Db   1492 AACACGCTGCATTTGGGTCACAGTTTATTTTTCCTGCTCTCCCTTGAGGACCATGTATAT 1551

Qy   1321 TTTGGTGAAATTTTTAAGAGATCAGAATTAGC-AATCACTCAATTTGGTTTTAAATTATC 1379
          |||| |||||||||||||||| |||||| ||| |||||||||||||||||||||||||||
Db   1552 TTTGATGAAATTTTTAAGAGACCAGAATCAGCAAATCACTCAATTTGGTTTTAAATTATC 1611

Qy   1380 CTGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAA--- 1436
          |||||||||||||||||| |||||||||||||||||||||||||||||||| | | |
Db   1612 CTGTATATATGTGATAACAAGCACTGGAAATAATTTATTCATCATTCTCATTTGTGAAGC 1671

Qy   1437 ACATTGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTACTTGTTTCTAC 1496
          | |||||||||||||||||||||||||||||||||| ||| |||||||| ||||||| ||
Db   1672 ATTCTGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTACTTGTTTCTAC 1731

Qy   1497 ATCAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGAC 1556
          |||||||||||||| ||||| |||||| || |||||||||| |||||||| |||| ||||
Db   1732 ATCAATCAGATCTTTCATCTACTTGGGAAATGAAAATACATATTCTAAAATATGAAGGAC 1791

Qy   1557 TTTTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATT 1616
          | |||||||||| |||| || | |||||||| |||||||||||||||||||| ||| |||
Db   1792 TCTTTGCCAAGTGTTATGAAATTTAGATTCTAACTTGTATAATGCTAACAAAGAAGAATT 1851

Qy   1617 ATTGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGAT 1676
          |||||| |||||||||||||||||| ||||||| || |||||||||| |||| ||||||
Db   1852 ATTGAACCACTTAATTCTTAATCTCTCTGACTTCAGAGCTGAACACTACTCACAATTGAT 1911

Qy   1677 AGAGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAG 1736
          |||||||| |||||||| |||||||| || | |||||||||||||||| || ||||||||
Db   1912 GTAGGTTTTACCTGTACCATGTTTGTGATGGTGGACTAATTGATGTAGTGAAGAAGAAAG 1971

Qy   1737 ATATGAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATG 1796
          | |||||||||||||||| |||||| |||| |||||||||||||||||||||||||||||
Db   1972 GCAGGAAGCAGAAATGTTAGTAGACTCCACACTCATTTTCATTAAGTTGTTTACAAATG 2031

Qy   1797 TAGATTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATT--TATAT 1854
          ||| || |||||||||||||||||||||| | | || || || |||||||  ||||| |
Db   2032 CAGA-TCTCAATATTTTAAGTTAAAGAAAGAAAGAAATACAATTTATGTGGTTGGGGGAG 2090

Qy   1855 GGTTTTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGA 1914
          | | || |   ||| | |||| |||  ||| | |||||||||||| |||||||||||||||
Db   2091 GCTGTTTGTTTTGTTTTGTTTTTTTTTTTTTTTTTTTGCCTAATAATACATGAGAGA 2150

Qy   1915 ATCTT 1919
          |||||
Db   2151 ATCTT 2155
```

```
RESULT 3
AF230460
LOCUS       AF230460                1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
  PUBMED    10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
     CDS             /note="cgFut9"
                     62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT       515 a    336 c    320 g    534 t
ORIGIN

Query Match           60.3%;  Score 1309.2;  DB 10;  Length 1705;
Best Local Similarity 88.9%;  Pred. No. 3.1e-212;
Matches 1520;  Conservative    0;  Mismatches  173;  Indels   17;  Gaps    9;

Qy      52 CACCGTGGAGTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAA 111
           ||| ||| ||| |||| ||| | |||||||| |||||| |||||||||| | ||||
Db       1 CACAGTGAAGTGGTTCTACACCCGTCTCTGCTGCCCATGC--TGTTCTCTACAGTAAAAA 58

Qy     112 ATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTG 171
           |||||||||||||||||||||||||||||| ||||| |||||||||||  | ||||||
Db      59 ATTATGACATCAACATCCAAAGGCATTCTTCGTCCATTCCTAATCGTCTGTGTTATCCTG 118

Qy     172 GGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGT 231
           | |||||||||||||||||| ||||| ||||||||||||||||||||||||||| ||||
Db     119 GCCTGCTTCATGGCATGTCTGCTTATTTACATCAAGCCCACCAACAGCTGGGTGTTCAGT 178

Qy     232 CCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGAT 291
           ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     179 CCAATGGAGTCTGCCAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGAT 238

Qy     292 TATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTT 351
           |||||||| ||||| || ||||| |||||||| |||||||||||||||||||||||||||
Db     239 TATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTT 298
```

```
Qy    352 ACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTG 411
Db    299 ACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACAGACCGCTCACTG 358

Qy    412 TACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAAC 471
Db    359 TACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAAC 418

Qy    472 TTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCC 531
Db    419 TTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAACCTAGAGTCTCCC 478

Qy    532 ACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGC 591
Db    479 ACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACTCTAACATATCGC 538

Qy    592 CGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTG 651
Db    539 CGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTG 598

Qy    652 TTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAG 711
Db    599 TTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAACTGGAATCCGGAG 658

Qy    712 CATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGC 771
Db    659 CATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATCCATACCTATGGG 718

Qy    772 CAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAA 831
Db    719 CAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATATCTACTTGTAAA 778

Qy    832 TTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAAT 891
Db    779 TTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAAAAGCTCTACAAT 838

Qy    892 GCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAAT 951
Db    839 GCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAAAACTATGAGAAT 898

Qy    952 TATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCA 1011
Db    899 TATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGTTGGCA 958

Qy   1012 AAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGA 1071
Db    959 AAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTATTTCAACTGGAGG 1018

Qy   1072 AAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGAC 1131
Db   1019 AAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGCCTGGCTTGTGAT 1078

Qy   1132 CATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAAT 1191
Db   1079 CACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAATGGTTTTGGAAT 1138

Qy   1192 TAAAGTGTCCATCATTGTCACAGT----GAAAAAGATTAATGAGACATAATTCAAGTTTT 1247
Db   1139 TAATGTTTCCCTCATTTTCAAATTTAAGGAAATGTATTAACATGACATCATATAAGTTGT 1198

Qy   1248 CAGGATAAGA--AGAAACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTT 1305
Db   1199 AAGGATAAGAAGAGAAACACCTTGCTTTCAAATCACAGTTTATTTTAATTGCCCTCCCTT 1258

Qy   1306 GAGGACCATGTATATTTTGGTG--AAATTTTTAAGAGATCAGAATTAGCAATCACTCAAT 1363
Db   1259 GAGTAACATGTATATTTTGATGGAAAATTTTTAAGAGCTCAGAATTAGCAATCACTTCAT 1318

Qy   1364 TTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCA 1423
Db   1319 TTGG-ATTAAATTATCCTGTATATACGTGATAATGAGCACTGAAAATAATTTATTCTCCA 1377

Qy   1424 CTCTCATTTATAAACATTGTTTTT--ACATTTTATAGTTGACTGTAAAGTAAATTTATGA 1481
Db   1378 CTCTCATTTGTAAACATTCTTTTTTCACATCTTATAGTTGTCTGTAAAGTAAGTTTATGA 1437

Qy   1482 TTTACTTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCC 1541
Db   1438 TTCAATTGTTTCTACATCAATCAGATCTTTAGTCTGTTTGGGAAATGGAAATGGGTGTCC 1497

Qy   1542 TAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGC 1601
Db   1498 TAAAATATGAAAG-CTTTTTGCAAAGTATTATAATGTCTAGATTCTAGTTTGCATAATGG 1556

Qy   1602 TAACAAAGAAGAATTATTGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACA 1661
Db   1557 TAACAAAGAAGTATTAGAGTATCACTTAATTCTTAACCTCTTTGAGTTCAGAGTTG-ACA 1615

Qy   1662 CTACTGACAACTGATAGAGTTTTTACTTG--TACCACATTTGTGAAGGGGAGCTAATTGA 1719
Db   1616 CTGTGTGTAATTGATGTACTTTTTAACTGTTTACCACATTTGTGAAGACAGACTGACTCA 1675

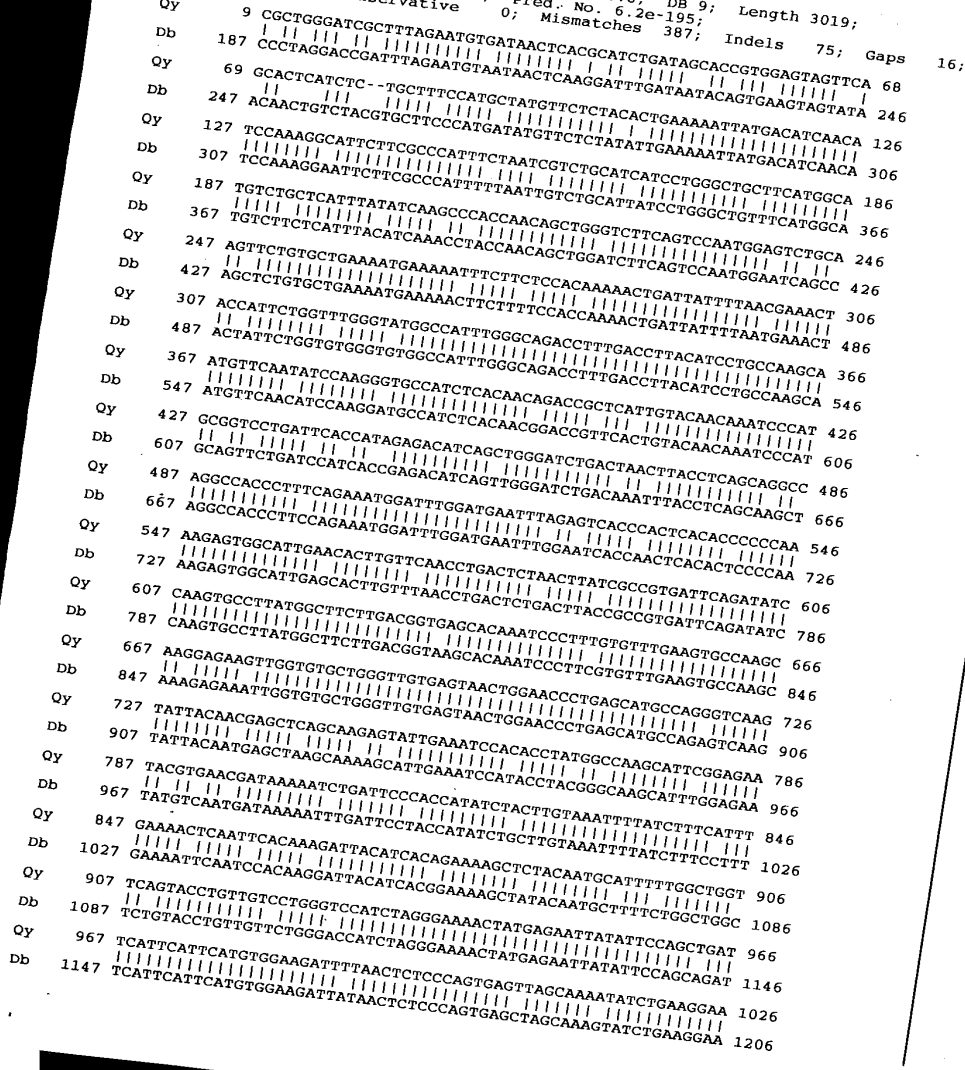
Qy   1720 TGTAGCAAATAAGAAAGATATGAAGCAGAA 1749
Db   1676 TGTTGTAAATAAGAAAGATATGAAGCAGAA 1705
```

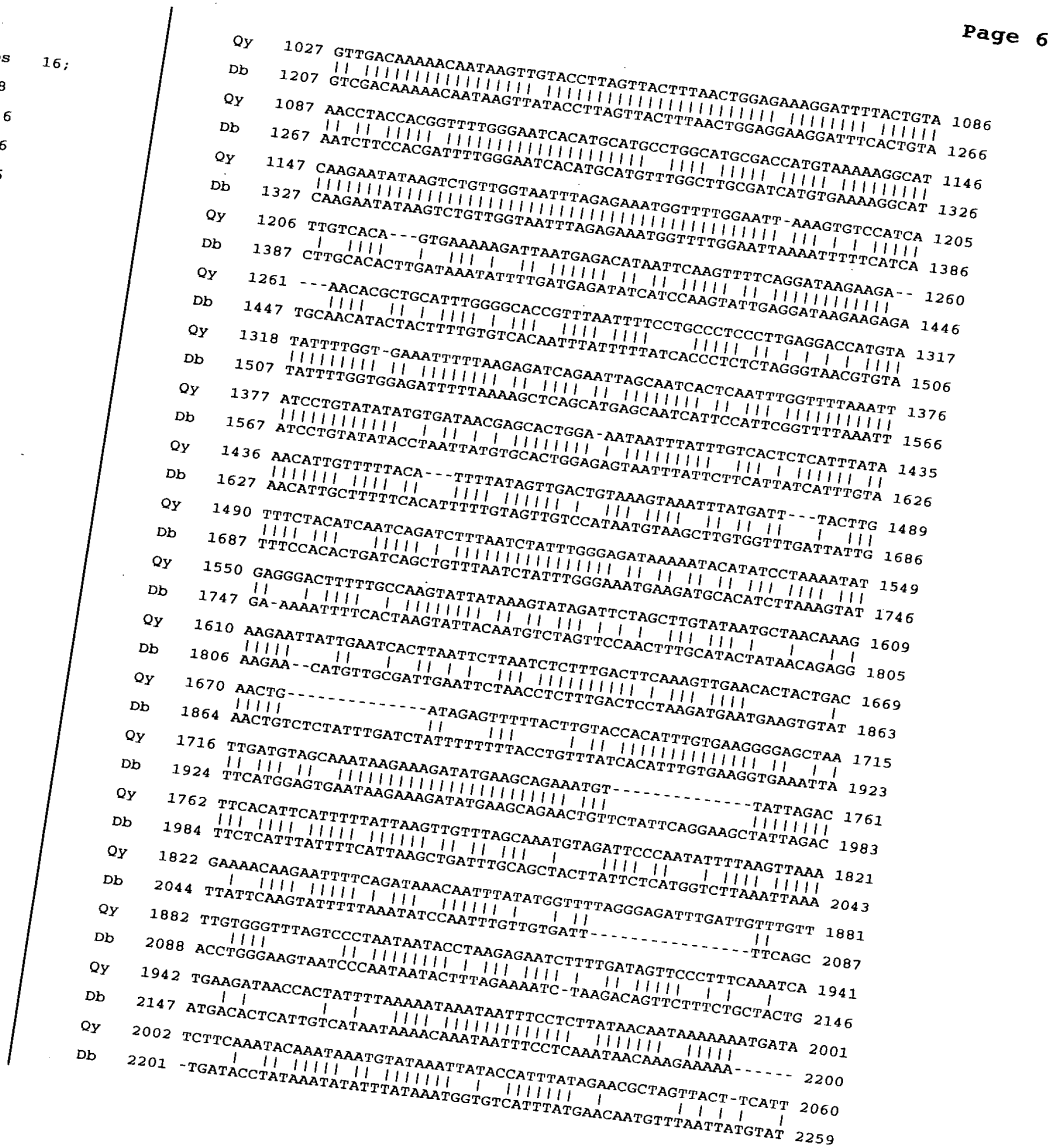
RESULT 4
AB023021

```
LOCUS       AB023021               3019 bp    mRNA    linear   PRI 23-JUN-1999
DEFINITION  Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB023021
VERSION     AB023021.1  GI:5139692
KEYWORDS    alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S.,
            Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and
            Narimatsu,H.
  TITLE     Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved
            between human and mouse; molecular cloning, characterization and
            tissue distribution of human Fuc-TIX
  JOURNAL   FEBS Lett. 453, 237-242 (1999)
REFERENCE   2  (bases 1 to 3019)
  AUTHORS   Kaneko,M., Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
```

ORIGIN

Query Match
Best Local Similarity 55.6%; Score 1206.8; DB 9; Length 3019;
Matches 1674; Conservative 78.4%; Pred. No. 6.2e-195;
0; Mismatches 387; Indels 75; Gaps 16;





```
Qy   2061 CATTTGAAATATTTTTTCCAAAGCATCCATATTTAA 2096
          || || ||   ||||||| ||||    ||||||||
Db   2260 CAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAA 2295


RESULT 5
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match           55.5%;  Score 1205;  DB 9;  Length 2501;
  Best Local Similarity   78.6%;  Pred. No. 1.3e-194;
  Matches 1668;  Conservative    0;  Mismatches  380;  Indels   75;  Gaps   16;

Qy     21 TTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGAGTAGTTCAGCACTCATCTC- 79
          |||||||||| |||||||| | || |||||  || ||| ||||||  | ||    |||
Db      9 TTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTAC 68

Qy     80 -TGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACATCCAAAGGCATT 138
           ||||| ||||| |||||||||||| | ||||||||||||||||||||||||||||| |||
Db     69 GTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATT 128

Qy    139 CTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATT 198
          |||||||||||| |||| |||||||| |||||||||| |||||||||||||| ||||||
Db    129 CTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATT 188

Qy    199 TATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTG 258
          || ||||| || |||||||||||| |||||||||||||||| || || || |||||||||
Db    189 TACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTG 248

Qy    259 AAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTT 318
          |||||||||| |||||  ||||| |||||||||||||||| ||||||||  ||||||||
Db    249 AAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTG 308

Qy    319 TGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATC 378
          |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    309 TGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATC 368

Qy    379 CAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATT 438
          |||||  |||||||||||||| |||||  ||| ||||||||||||||||||| || |||||
Db    369 CAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATC 428

Qy    439 CACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTT 498
          || ||  ||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db    429 CATCACCGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAAGCTAGGCCACCCTTC 488

Qy    499 CAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATT 558
          |||||||||||||||||||||| || ||||| |||||||| ||||||||||||||||||
Db    489 CAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATT 548

Qy    559 GAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTAT 618
          || |||||||| |||||||||| |||||||||||||||||||||||||||||||||||||
Db    549 GAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTAT 608

Qy    619 GGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTG 678
          ||||||||||||||| |||||||||||||| |||||||||||||||||||| ||||| |||
Db    609 GGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTG 668

Qy    679 GTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAG 738
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db    669 GTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAG 728

Qy    739 CTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGAT 798
          || ||||| || |||||||||| ||||| || |||||||| |||||||| || || |||
Db    729 CTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGAT 788

Qy    799 AAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATT 858
          |||||| ||||||| ||||||||| |||||||||||||||||| |||||||| |||||
Db    789 AAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATC 848

Qy    859 CACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTT 918
          ||||| |||||||||| |||||||| ||||||| ||| ||||||| || |||||||||
Db    849 CACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTT 908

Qy    919 GTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCAT 978
          || ||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    909 GTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCAT 968

Qy    979 GTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAAC 1038
          ||||||||| |||||||||||||||| |||||| ||||||||||||||| ||||||||||
Db    969 GTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAAC 1028

Qy   1039 AATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGG 1098
          |||||||| |||||||||||||||||||| |||||||| |||||||| || |||||
Db   1029 AATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGA 1088

Qy   1099 TTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAG 1158
          |||||||||||||||||||  |||| ||||| ||||| ||||||||||||||||||||||
Db   1089 TTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAG 1148
```

```
Qy   1159 TCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT-AAAGTGTCCATCATTGTCACA---G 1214
          ||||||||||||||||||||||||||||||||||| ||| | | ||||| | ||||    |
Db   1149 TCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTG 1208

Qy   1215 TGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA-----AACACGCTG 1269
          ||| |  || |||||| || || ||||| || ||||||||||||     ||||  ||
Db   1209 ATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTA 1268

Qy   1270 CATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGT-GA 1328
          | |||| | |||  |||| ||||      ||||| || | | | | ||||||||||||| ||
Db   1269 CTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGA 1328

Qy   1329 AATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATA 1388
          |||||||| || |||| || |||||||| || ||| ||||||||||||||||||||||||
Db   1329 GATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATA 1388

Qy   1389 TGTGATAACGAGCACTGGA-AATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTT 1447
          | || | | |||||||| | |||||||||  ||| | |||||  ||||||||| ||||
Db   1389 CCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTT 1448

Qy   1448 ACA---TTTTATAGTTGACTGTAAAGTAAATTTATGATT---TACTTGTTTCTACATCAA 1501
          |||   |||| ||||||| |  ||| ||||  || || ||   | |||||| |||    |
Db   1449 TCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGA 1508

Qy   1502 TCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTTT 1561
          |||| | |||||||||||||||| || || || || ||| |||| |||||    | ||||
Db   1509 TCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGA-AAAATTTTC 1567

Qy   1562 GCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTGA 1621
          |  |||||||| || || ||| | | |||| ||| |  |  |  | ||||||    ||||
Db   1568 ACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAA--CATGT 1625

Qy   1622 ATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTG------- 1674
          |  || |  ||| |||||||||| | ||| ||||  |  |||||
Db   1626 TGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTGTATAACTGTCTCTAT 1685

Qy   1675 ------ATAGAGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAA 1728
                || |||  |  || |||||||||||||| || | ||| ||| ||  ||
Db   1686 TGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAATTATTCATGGAGTGAA 1745

Qy   1729 TAAGAAAGATATGAAGCAGAAATGT-------------TATTAGACTTCACATTCATTT 1774
          |||||||||||||||||||||||||             |||||||||| |||| ||||
Db   1746 TAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTAGACTTCTCATTTATTT 1805

Qy   1775 TTATTAAGTTGTTTAGCAAATGTAGATTCCCAATATTTAAGTTAAAGAAAACAAGAATT 1834
          | |||||| || || |||  |  |||| ||    |||| |||||  |   |||| |||
Db   1806 TCATTAAGCTGATTTGCAGCTACTTATTCTCATGGCCTTAAATTAAATTATTCAAGTATT 1865

Qy   1835 TTCAGATAAACAATTTATATGGTTTTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGT 1894
          || | |||  |||||||  | |||  |  |||  |  || |
Db   1866 TTTAAATATCCAATTTGTTGTGATTTTCAGCACCTGGAAGTAAT--------------- 1909

Qy   1895 CCCTAATAATACCTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCATGAAGATAACCAC 1954
          || |||||||| | ||| |||| |  || ||||| | | |  |  |  |  |||||||
Db   1910 -CCCAATAATACTTTAGAAAATC-TAAGACAGTTCTTTCTGCTACTGATGACACTCATTG 1967

Qy   1955 TATTTTAAAAATAAATAATTTCCTCTTATAACAATAAAAAAATGATATCTTCAAATACAA 2014
          | |  |||| ||||||||||||||  |||||||  |||||                  |
Db   1968 TCATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAA-------TGATACCTATAA 2020

Qy   2015 ATAAATGTATAAATTATACCATTTATAGAACGCTAGTTACT-TCATTCATTTGAAATATT 2073
          ||| || |||||||  |  |||||||  |  | | | |||| || ||  ||||||||||
Db   2021 ATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTAAGATTT 2080

Qy   2074 TTTTCCAAAGCATCCATATTTAA 2096
          |||||  ||||  ||||||||
Db   2081 TTTTCTGAAGCCCTAATATTTAA 2103

RESULT 6
AL512406
LOCUS       AL512406                 80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-504J9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-381A2 is at 80148 in this sequence.
            The true right end of clone RP11-77B15 is at 100 in this sequence.
FEATURES    Location/Qualifiers
     source          1. .80247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-504J9"
                     /clone_lib="RPCI-11.2"
     repeat_region   184. .1422
                     /note="L2 repeat: matches 875. .2419 of consensus"
     repeat_region   1744. .1834
                     /note="MIR repeat: matches 166. .258 of consensus"
     repeat_region   2950. .2977
                     /note="14 copies 2 mer ca 100% conserved"
     repeat_region   3411. .3669
                     /note="LTR16C repeat: matches 93. .364 of consensus"
     repeat_region   4372. .4704
                     /note="MLT1A1 repeat: matches 5. .365 of consensus"
     misc_feature    complement(7821. .8406)
                     /note="match: GSS: Em:AQ550748"
     repeat_region   8126. .8184
                     /note="MER58B repeat: matches 272. .332 of consensus"
     repeat_region   8185. .8696
                     /note="MLT2D repeat: matches 1. .513 of consensus"
     repeat_region   8779. .8817
                     /note="MLT2D repeat: matches 504. .542 of consensus"
     repeat_region   8819. .8974
                     /note="MER58A repeat: matches 6. .162 of consensus"
```

```
misc_feature    complement(9443. .9891)
                /note="match: GSS: Em:B90633"
misc_feature    complement(9452. .9883)
                /note="match: GSS: Em:AQ119695"
repeat_region   9782. .9894
                /note="MIR repeat: matches 92. .205 of consensus"
repeat_region   10089. .10212
                /note="MIR repeat: matches 135. .262 of consensus"
misc_feature    complement(10093. .10671)
                /note="match: GSS: Em:AQ237061"
repeat_region   10223. .10282
                /note="MIR repeat: matches 203. .262 of consensus"
misc_feature    complement(10315. .10631)
                /note="match: GSS: Em:AQ200259"
misc_feature    10805. .11186
                /note="match: GSS: Em:AQ072353"
repeat_region   11657. .11844
                /note="L2 repeat: matches 2546. .2750 of consensus"
misc_feature    complement(11659. .12084)
                /note="match: GSS: Em:AQ518361"
misc_feature    complement(11674. .11987)
                /note="match: GSS: Em:AQ094678"
misc_feature    12083. .12588
                /note="match: GSS: Em:AQ691235"
repeat_region   12596. .12810
                /note="MIR repeat: matches 23. .248 of consensus"
repeat_region   12920. .12986
                /note="L1MA5 repeat: matches 6228. .6295 of consensus"
repeat_region   12987. .13169
                /note="L1MA5 repeat: matches 5852. .6043 of consensus"
repeat_region   13168. .13280
                /note="L1MA8 repeat: matches 5710. .5832 of consensus"
repeat_region   13851. .17218
                /note="L1PB1 repeat: matches 2787. .6155 of consensus"
repeat_region   18903. .18953
                /note="MLT2CB repeat: matches 439. .501 of consensus"
repeat_region   18984. .19433
                /note="MLT2CB repeat: matches 1. .460 of consensus"
misc_feature    complement(19108. .19690)
                /note="match: GSS: Em:AQ563935"
repeat_region   20683. .20801
                /note="AluJo/FRAM repeat: matches 169. .288 of consensus"
repeat_region   20802. .20862
                /note="AluJ/FLAM repeat: matches 3. .66 of consensus"
repeat_region   20912. .21033
                /note="MIR repeat: matches 114. .237 of consensus"
repeat_region   21047. .21128
                /note="41 copies 2 mer ta 63% conserved"
repeat_region   21676. .21859
                /note="L2 repeat: matches 2130. .2308 of consensus"
repeat_region   21890. .23612
                /note="L1P repeat: matches 841. .2570 of consensus"
repeat_region   23601. .26609
                /note="L1PA5 repeat: matches 3107. .6145 of consensus"
repeat_region   26620. .26789
                /note="MER5A repeat: matches 1. .174 of consensus"
repeat_region   28521. .28627
                /note="MER39 repeat: matches 5. .111 of consensus"
repeat_region   28656. .28837
                /note="MER39b repeat: matches 54. .235 of consensus"
repeat_region   28838. .29142
                /note="AluSc repeat: matches 4. .308 of consensus"
repeat_region   29143. .29494
                /note="MER39b repeat: matches 235. .549 of consensus"
repeat_region   30529. .31344
                /note="L1PA3 repeat: matches 5331. .6146 of consensus"
repeat_region   32234. .32340
                /note="L2 repeat: matches 2627. .2750 of consensus"
repeat_region   32399. .32578
                /note="MIR repeat: matches 12. .203 of consensus"
repeat_region   32816. .32979
                /note="MIR repeat: matches 91. .259 of consensus"
misc_feature    complement(33006. .33450)
                /note="match: GSS: Em:AQ261227"
misc_feature    complement(33058. .33444)
                /note="match: GSS: Em:AQ040453"
repeat_region   33312. .33596
                /note="L1M4 repeat: matches 3473. .3773 of consensus"
repeat_region   34856. .34915
                /note="30 copies 2 mer aa 68% conserved"
misc_feature    complement(34987. .35547)
                /note="match: GSS: Em:AQ321598
                match: STS: Em:G55571"
misc_feature    complement(36488. .36936)
                /note="match: GSS: Em:B41073"
misc_feature    complement(36654. .37090)
                /note="match: GSS: Em:AQ772584"
misc_feature    37102. .37591
                /note="match: GSS: Em:AQ489567"
misc_feature    complement(37581. .38221)
                /note="match: GSS: Em:AQ418233"
repeat_region   37720. .37772
                /note="L2 repeat: matches 2657. .2710 of consensus"
misc_feature    complement(37778. .38215)
                /note="match: GSS: Em:AQ819773"
misc_feature    38218. .38732
                /note="match: GSS: Em:AQ599868"
repeat_region   38861. .39164
                /note="AluSx repeat: matches 1. .305 of consensus"
repeat_region   39612. .39653
                /note="21 copies 2 mer ac 90% conserved"
repeat_region   40854. .40909
                /note="L2 repeat: matches 2655. .2710 of consensus"
repeat_region   42169. .42270
                /note="MER5A repeat: matches 1. .111 of consensus"
repeat_region   42411. .42548
                /note="MER5B repeat: matches 21. .181 of consensus"
repeat_region   42780. .42938
                /note="MER5B repeat: matches 17. .177 of consensus"
repeat_region   43229. .43326
                /note="MIR repeat: matches 40. .148 of consensus"
repeat_region   45383. .45428
                /note="MER5B repeat: matches 18. .62 of consensus"
misc_feature    complement(47284. .47971)
                /note="match: STS: Em:G51995
                match: GSS: Em:AQ269059"
misc_feature    complement(47537. .48011)
                /note="match: STS: Em:G36899"
misc_feature    complement(47723. .48014)
                /note="match: STS: Em:G30049"
misc_feature    47972. .48451
                /note="match: GSS: Em:AQ336527"
misc_feature    complement(48518. .48985)
                /note="match: GSS: Em:AQ801834"
repeat_region   48597. .48803
```

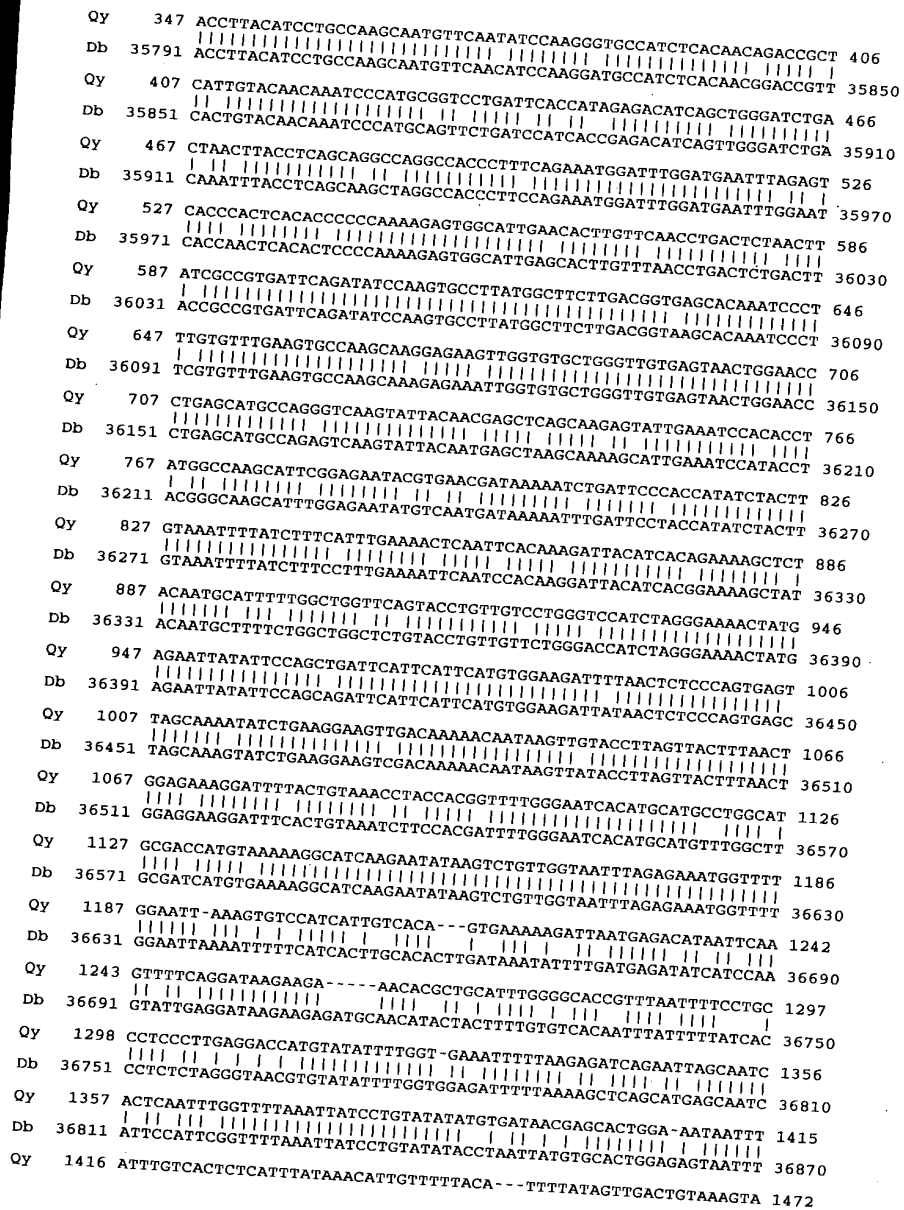
Query Match 53.6%; Score 1162.4; DB 9; Length 80247;
Best Local Similarity 78.7%; Pred. No. 1.4e-187;
Matches 1602; Conservative 0; Mismatches 361; Indels 73; Gaps 15;

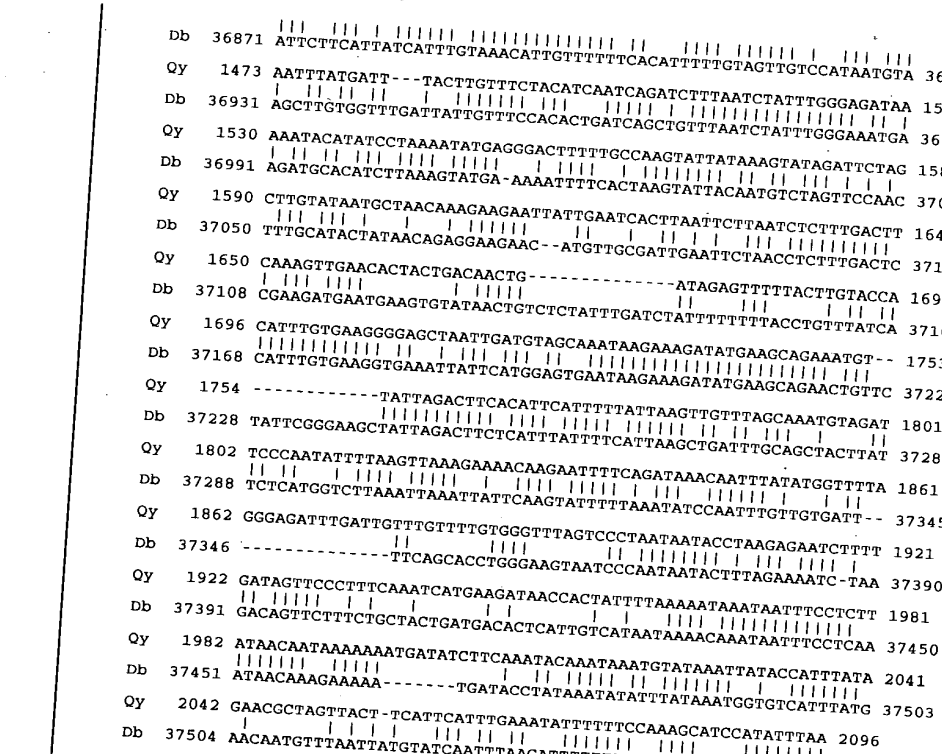
```
Qy    107 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 166
          |||||||||||||||||||||||||||| |||||||||||||| |||| |||||||| |
Db  35551 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 35610

Qy    167 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 226
          |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db  35611 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 35670

Qy    227 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 286
          |||||||||||| || || || ||||||||||||||||||||| ||||| ||||| ||||
Db  35671 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 35730

Qy    287 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 346
          |||||||||||| |||||||| ||||| |||||||||||||||||||||||||||||||
Db  35731 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 35790
```





RESULT 7
AF345993
LOCUS AF345993 1128 bp mRNA linear ROD 07-MAR-2001
DEFINITION Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA, complete cds.
ACCESSION AF345993
VERSION AF345993.1 GI:13242183
KEYWORDS .
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1128)
AUTHORS Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
TITLE Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX) and comparison of the expression of fuc-TIV and fuc-TIX genes during rat postnatal cerebellum development
JOURNAL J. Neurosci. Res. 62 (2), 206-215 (2000)
PUBMED 11020213
REFERENCE 2 (bases 1 to 1128)
AUTHORS Smith,F.I. and Baboval,T.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver Center, 200 Trapelo Rd., Waltham, MA, USA

```
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT       324 a    256 c    228 g    320 t
ORIGIN
```

Query Match 48.4%; Score 1049.6; DB 10; Length 1128;
Best Local Similarity 95.7%; Pred. No. 2.9e-168;
Matches 1079; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy      67 CAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACA 126
           |||||| |||||||||||||||| |||||||||||  ||||||||||||| ||||||
Db       1 CAGCACGCATCTCTGCTTTCCATGATATGTTCTCTACCGTGAAAAATTATGACCTCAACA 60

Qy     127 TCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCA 186
           |||||||||||||||||||||||||||||||||||| |||||||||||||| |||||
Db      61 TCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGGGCTGCTTCGTGGCA 120

Qy     187 TGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCA 246
           ||||| || || || ||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCA 180

Qy     247 AGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACT 306
           || ||||||||||||||||||||||||||||||  |||||||||||||||| ||||||
Db     181 AGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATTATTTTAATGAAACT 240

Qy     307 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 366
           || |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     241 ACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCC 300

Qy     367 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426
           ||||||||| ||||||||||||||||||||||| ||||||| |||||||||||||||||
Db     301 ATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGTACAACAAATCCCAT 360

Qy     427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 486
           |||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 420

Qy     487 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 546
           ||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     421 AGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCACTCACACCCCCCAA 480

Qy     547 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 606
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATA 540

Qy     607 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 666
           |||||||||| |||||||||||||||||||  |||||||||||||||||||||||||||
Db     541 CAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTTGAAGTGCCAAGC 600

Qy     667 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 726
           || |||||||| |||||||||||||||||||||||| || ||||||||||||||||||||
Db     601 AAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCATGCCAGGGTCAAG 660

Qy     727 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 786
           |||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||
Db     661 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGCAAGCATTTGGAGAA 720

Qy     787 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 846
           || ||||| |||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     721 TATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCCTTT 780

Qy     847 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 840

Qy     907 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 966
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 900

Qy     967 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 1026
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     901 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTGGCAAAATATCTGAAGGAG 960

Qy    1027 GTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTA 1086
           |||||||||||||||||||| || |||||||||| |||||||| ||||||||||||||||
Db     961 GTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGAAGGATTTTACTGTA 1020

Qy    1087 AACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCAT 1146
           ||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     1021 AACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATCATGTAAAAAGGCAT 1080

Qy    1147 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     1081 CAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1128
```

```
RESULT 8
AC100674/c
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
```

```
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L15932
    Center clone name: 167_K_24
--------------.
* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      788: contig of 788 bp in length
*      789 888: gap of      100 bp
*      889     1702: contig of 814 bp in length
*     1703 1802: gap of      100 bp
*     1803     2642: contig of 840 bp in length
*     2643 2742: gap of      100 bp
*     2743     3637: contig of 895 bp in length
*     3638 3737: gap of      100 bp
*     3738     4635: contig of 898 bp in length
*     4636 4735: gap of      100 bp
*     4736     5657: contig of 922 bp in length
*     5658 5757: gap of      100 bp
*     5758     6692: contig of 935 bp in length
*     6693 6792: gap of      100 bp
*     6793     7703: contig of 911 bp in length
*     7704 7803: gap of      100 bp
*     7804     8742: contig of 939 bp in length
*     8743 8842: gap of      100 bp
*     8843     9787: contig of 945 bp in length
*     9788 9887: gap of      100 bp
*     9888    10819: contig of 932 bp in length
*    10820 10919: gap of      100 bp
*    10920    11808: contig of 889 bp in length
*    11809 11908: gap of      100 bp
*    11909    12830: contig of 922 bp in length
*    12831 12930: gap of      100 bp
*    12931    13847: contig of 917 bp in length
*    13848 13947: gap of      100 bp
*    13948    14886: contig of 939 bp in length
*    14887 14986: gap of      100 bp
*    14987    15911: contig of 925 bp in length
*    15912 16011: gap of      100 bp
*    16012    16903: contig of 892 bp in length
*    16904 17003: gap of      100 bp
*    17004    17906: contig of 903 bp in length
*    17907 18006: gap of      100 bp
*    18007    18931: contig of 925 bp in length
*    18932 19031: gap of      100 bp
*    19032    19945: contig of 914 bp in length
*    19946 20045: gap of      100 bp
*    20046    20951: contig of 906 bp in length
*    20952 21051: gap of      100 bp
*    21052    21949: contig of 898 bp in length
*    21950 22049: gap of      100 bp
*    22050    22950: contig of 901 bp in length
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
```

```
            *    62400    63322: contig of 923 bp in length
            *    63323 63422: gap of      100 bp
            *    63423    64336: contig of 914 bp in length
            *    64337 64436: gap of      100 bp
            *    64437    65354: contig of 918 bp in length
            *    65355 65454: gap of      100 bp
            *    65455    66352: contig of 898 bp in length
            *    66353 66452: gap of      100 bp
            *    66453    67362: contig of 910 bp in length
            *    67363 67462: gap of      100 bp
            *    67463    68370: contig of 908 bp in length
            *    68371 68470: gap of      100 bp
            *    68471    69379: contig of 909 bp in length
            *    69380 69479: gap of      100 bp

  Query Match             40.7%;  Score 883;  DB 2;  Length 71396;
  Best Local Similarity   88.0%;  Pred. No. 2.7e-140;
  Matches  906;  Conservative    0;  Mismatches  122;  Indels    1;  Gaps    1;

Qy     922 CTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTG 981
           | | | | |   | |
Db   30210 CAGTGGCGAGACATCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 30151

Qy     982 GAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAAT 1041
                                                                   |||
Db   30150 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAT 30091

Qy    1042 AAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCAC-GGTT 1100
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||| |||
Db   30090 AAGTNGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGGTT 30031

Qy    1101 TTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTC 1160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30030 TTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTC 29971

Qy    1161 TGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTCCATCATTGTCACAGTGAAAA 1220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29970 TGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTCCATCATTGTCACAGTGAAAA 29911

Qy    1221 AGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGAAACACGCTGCATTTGGGGCA 1280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29910 AGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGAAACACGCTGCATTTGGGGCA 29851

Qy    1281 CCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAGAG 1340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29850 CCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAGAG 29791

Qy    1341 ATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAG 1400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29790 ATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAG 29731

Qy    1401 CACTGGAAATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAGTT 1460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29730 CACTGGAAATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAGTT 29671

Qy    1461 GACTGTAAAGTAAATTTATGATTTACTTGTTTCTACATCAATCAGATCTTTAATCTATTT 1520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29670 GACTGTAAAGTAAATTTATGATTTACTTGTTTCTACATCAATCAGATCTTTAATCTATTT 29611

Qy    1521 GGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTAT 1580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29610 GGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTAT 29551

Qy    1581 AGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTGAATCACTTAATTCTTAATCT 1640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29550 AGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTGAATCACTTAATTCTTAATCT 29491

Qy    1641 CTTTGACTTCAAAGTTGAACACTACTGACAACTGATAGAGTTTTTACTTGTACCACATTT 1700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29490 CTTTGACTTCAAAGTTGAACACTACTGACAACTGATAGAGTTTTTACTTGTACCACATTT 29431

Qy    1701 GTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATATGAAGCAGAAATGTTATTAGA 1760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   29430 GTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATATGAAGCAGAAATGTTATTAGA 29371

Qy    1761 CTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGATTCCCAATATTTTAAGTTAA 1820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29370 CTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGATTCCCAATATTTTAAGTTAA 29311

Qy    1821 AGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTTAGGGAGATTTGATTGTTTGT 1880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29310 AGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTTAGGGAGATTTGATTGTTTGT 29251

Qy    1881 TTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTTTGATAGTTCCCTTTCAAATC 1940
           ||||||||||||||||||||||||||||||||||||||||||||||||    | ||   |
Db   29250 TTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTTTGATAGTTCGAATCCACCAC 29191

Qy    1941 ATGAAGATA 1949
           | |  || |
Db   29190 AGGGCGAGA 29182


RESULT 9
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match             34.1%;  Score 740;  DB 5;  Length 1707;
  Best Local Similarity   72.2%;  Pred. No. 6.9e-116;
  Matches 1002;  Conservative    1;  Mismatches  356;  Indels   29;  Gaps    2;

Qy     108 AAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCAT 167
           | |||||||||||||||||| ||||| ||| | || |||||| | || |||| |||  |
Db     261 ACAAATTATGACATCAACATCTAAAGGAATTTTCCGGCCATTTTTTATTGTCTTCATTGT 320
```

```
Qy    168 CCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTT 227
Db    321 YCTTGGTTGTTTCATGGCATTTATACTAATTTATATTAAACCAACAAGCAGCTGGATCTC 380
Qy    228 CAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAAC 287
Db    381 TAATCCTGTGGAATCAGCCAGCTCAGTTTTGAAAATGAAGAGCTTCTTTTCTTCCAAAAC 440
Qy    288 TGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGA 347
Db    441 TGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGACATTCGA 500
Qy    348 CCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTC 407
Db    501 TCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTGACCGCTC 560
Qy    408 ATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGAC 467
Db    561 ACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGGATCTGAC 620
Qy    468 TAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTC 527
Db    621 TAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACTTGGAATC 680
Qy    528 ACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTA 587
Db    681 TCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCCTGACTTA 740
Qy    588 TCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTT 647
Db    741 CCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAGGCGCCTT 800
Qy    648 TGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCC 707
Db    801 TACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACTGGAACCC 860
Qy    708 TGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTA 767
Db    861 TGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCCATACCTA 920
Qy    768 TGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTG 827
Db    921 CGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCTCCAGTTG 980
Qy    828 TAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTA 887
Db    981 CAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGAAACTTTA 1040
Qy    888 CAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGA 947
Db   1041 CAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAAACTACGA 1100
Qy    948 GAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTT 1007
Db   1101 GAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAAGAGAGCT 1160
Qy   1008 AGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTG 1067
Db   1161 TGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACTTCAACTG 1220
Qy   1068 GAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATG 1127
Db   1221 GAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTCTTGCTTG 1280
Qy   1128 CGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTG 1187
Db   1281 TGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAATGGTTTTG 1340
Qy   1188 GAATTAA------AGTGTCCATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCA 1241
Db   1341 GAACTAATTTGTGTGTGTGTATGTGCACTTTTTTGAAGAAGATAATGGATCTTTAGTCTA 1400

Qy   1242 AGTTTTCAGGATAAGAAGAAACACGCTGCA----------TTTGGGG 1278
Db   1401 AGTGGAGAGTAAAGGAATAAGGAAAAAAGAAGGAAGGATGAACCTCACAGAACCTTACGT 1460
Qy   1279 CACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAG 1338
Db   1461 CAAGGTTTATCAGCCATCCTCTCTTGGCTATTCTATTTATATTTTGTAAGAGATTTTAAA 1520
Qy   1339 AGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACG 1398
Db   1521 AATTATGAGCAGCAGTCATCAAAACTCAGTTTTAAATTATAATGTACATACTAATTATGT 1580
Qy   1399 AGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAG 1458
Db   1581 GCACTGAAGAATATATGATTGTTTACTATAATTTTTAAACAAGATTTTTCACATTTTCTG 1640
Qy   1459 TTGACTGT 1466
Db   1641 TGGAGTAT 1648
```

RESULT 10
AC100674
LOCUS AC100674 71396 bp DNA linear HTG 22-NOV-2001
DEFINITION Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC100674
VERSION AC100674.1 GI:17048040
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 71396)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-167K24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 71396)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit edu
---------- Project Information
Center project name L15932
Center clone name 167_K_24

```
--------------.
* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      788: contig of 788 bp in length
*      789 888: gap of      100 bp
*      889     1702: contig of 814 bp in length
*     1703 1802: gap of      100 bp
*     1803     2642: contig of 840 bp in length
*     2643 2742: gap of      100 bp
*     2743     3637: contig of 895 bp in length
*     3638 3737: gap of      100 bp
*     3738     4635: contig of 898 bp in length
*     4636 4735: gap of      100 bp
*     4736     5657: contig of 922 bp in length
*     5658 5757: gap of      100 bp
*     5758     6692: contig of 935 bp in length
*     6693 6792: gap of      100 bp
*     6793     7703: contig of 911 bp in length
*     7704 7803: gap of      100 bp
*     7804     8742: contig of 939 bp in length
*     8743 8842: gap of      100 bp
*     8843     9787: contig of 945 bp in length
*     9788 9887: gap of      100 bp
*     9888    10819: contig of 932 bp in length
*    10820 10919: gap of      100 bp
*    10920    11808: contig of 889 bp in length
*    11809 11908: gap of      100 bp
*    11909    12830: contig of 922 bp in length
*    12831 12930: gap of      100 bp
*    12931    13847: contig of 917 bp in length
*    13848 13947: gap of      100 bp
*    13948    14886: contig of 939 bp in length
*    14887 14986: gap of      100 bp
*    14987    15911: contig of 925 bp in length
*    15912 16011: gap of      100 bp
*    16012    16903: contig of 892 bp in length
*    16904 17003: gap of      100 bp
*    17004    17906: contig of 903 bp in length
*    17907 18006: gap of      100 bp
*    18007    18931: contig of 925 bp in length
*    18932 19031: gap of      100 bp
*    19032    19945: contig of 914 bp in length
*    19946 20045: gap of      100 bp
*    20046    20951: contig of 906 bp in length
*    20952 21051: gap of      100 bp
*    21052    21949: contig of 898 bp in length
*    21950 22049: gap of      100 bp
*    22050    22950: contig of 901 bp in length
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
```

```
                *    67463    68370: contig of 908 bp in length
                *    68371 68470: gap of     100 bp
                *    68471    69379: contig of 909 bp in length
                *    69380 69479: gap of     100 bp

  Query Match                34.1%;  Score 740;  DB 2;  Length 71396;
  Best Local Similarity      83.4%;  Pred. No. 4.3e-116;
  Matches  844;  Conservative    0;  Mismatches  154;  Indels   14;  Gaps    6;

Qy    186 ATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGC 245
          || |||   | | |    | ||||||||||||| ||||||||||||||||||||||||||||
Db  65456 ATCTCTCGCCCTGTGGTGGATTCCCACCAACAGCTCGGTCTTCAGTCCAATGGAGTCTGC 65515

Qy    246 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65516 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 65575

Qy    306 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65576 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 65635

Qy    366 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65636 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 65695

Qy    426 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65696 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 65755

Qy    486 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65756 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 65815

Qy    546 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65816 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 65875

Qy    606 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65876 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 65935

Qy    666 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65936 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 65995

Qy    726 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65996 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 66055

Qy    786 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATT 845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  66056 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTC-TC 66114

Qy    846 TGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGG 905
          |||||||||| |  |  |  |  ||  | ||  |  |||||||||| |||||||||||
Db  66115 TGAAAACTCATTCACAAGATTACTCCCAAAAG--------TCCACATGCTTTTTGGCTGG 66166

Qy    906 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66167 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 66226

Qy    966 TTCATTCATTC--ATGTGGAAGATTTTAACTCT-CCCAGTGAGTTAGCAAAATATCTGAA 1022
          |||||||||||  |||||||||| ||||| |||||||||||||||||  |||||  ||||
Db  66227 TTCATTCATTCCATCGTGGAAGATTTTCACTCTCCCCAGTGAGTTAGCAAATATCCTGAA 66286

Qy    1023 GGAAGTTGACAAAAACAATAAGTT-GTACCTTAGTTACTTTAACT-GGAGAAAGGATTTT 1080
           |||||| ||||| |||||||||||  |||||||||||||||||| | |||||||||||| |
Db  66287 GGAAGTCGACAANAACAATAAGTTGGTACCTTAGTTACTCTAACTCGNAGAAAGGATTCT 66346

Qy    1081 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1140
           || |
Db  66347 ACCGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 66406

Qy    1141 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1192

Db  66407 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGATT 66458

RESULT 11
AB035905
LOCUS
DEFINITION  AB035905                1080 bp    DNA     linear   VRT 05-JAN-20
ACCESSION   Xenopus laevis gene for xFUT9, complete cds.
VERSION     AB035905
KEYWORDS    AB035905.1  GI:18146863
SOURCE      xFUT9.
  ORGANISM  Xenopus laevis DNA.
            Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
     gene            /db_xref="taxon:8355"
                     1. .1080
     CDS             /gene="xFuc-TIX"
                     1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT       332 a    231 c    202 g    315 t
ORIGIN

  Query Match                31.1%;  Score 675.2;  DB 5;  Length 1080;
  Best Local Similarity      76.6%;  Pred. No. 6.7e-105;
  Matches  827;  Conservative    0;  Mismatches  253;  Indels    0;  Gaps    0;

Qy    115 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 174
          |||||||| |||| ||||| ||  | ||  ||||| ||   ||| ||    |  ||  |
Db      1 ATGACATCAGCATCTAAAGGAATCTTACGGGCATTTTTAGCTGTCAGCGGTGTGTTGATC 60

Qy    175 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 234
          || || |||  |||| | || ||  |||| || || |||| |||| | |  ||||||
Db     61 TGTTTTATGATATGTTTACTGATATACGTCAAACCAACAAACAACTGGATTTCCAGTCCT 120

Qy    235 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 294
          || || || |||   ||| || || |||||||| ||||| |||||| | |||    |||||
Db    121 ATCGAATCAGCACATTCAGTCCTTAAAATGAAGAATTTTTTCTCCTCTAAACAGAATTAT 180

Qy    295 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 354
          |  || ||||   | |||| |||||| ||||| |||||||||| ||||| || | |
Db    181 TACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGACTTTTGAACTGAAA 240

Qy    355 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 414
          |||||||  || |||| ||||||  | |||||||||||| || || ||||||| || | |||

```

```
Db    241 TCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTGACCGCGCACTCTAC 300

Qy    415 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 474
          ||||| || ||||| ||||| |||||| |||||||| |||||||||| | |||| |
Db    301 AACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGGATCTCATTAACCTG 360

Qy    475 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 534
          || || |||   |||||||| || |||||||||||||||||||| | || ||||| ||
Db    361 CCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATCTGGAATCACCAACA 420

Qy    535 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 594
          || || || ||||| |||||||||| ||||| || |||||||||| ||||| | |||
Db    421 CATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTCTGACTTACAGGCGT 480

Qy    595 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 654
          || ||||||||||||||||||||||||||| ||   || |||||||| || |||| |||
Db    481 GACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAAAACCATTTGACTTT 540

Qy    655 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 714
          || |||||||||| || ||||| || || |||||||| || |||||||||| || |||
Db    541 GAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACTGGAACCCAGAACAT 600

Qy    715 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 774
          ||||| ||||| |||||||| || || | |||    ||||||||    || ||||| |||
Db    601 GCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTATAACTTATGGGCAA 660

Qy    775 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 834
          || || || ||||| ||| |||||||| ||| |||| || ||  |  |||| ||||||
Db    661 GCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTGCATCTTGCAAATTT 720

Qy    835 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 894
          ||||| || |||||||| || |||||||||||||| |||||||||| || || ||||||
Db    721 TATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAAAACTTTATAATGCA 780

Qy    895 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 954
          || |||||||| ||||| ||  ||||  |||| ||  |||| ||||||||||||||| ||
Db    781 TTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAAACTATGAGAACTAC 840

Qy    955 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 1014
          ||||| || ||||| || || || |||||||||||||  |||||| || || |||||  |
Db    841 ATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAAGAGAATTAGCTGAC 900

Qy   1015 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
           ||||       |  ||||| | |   |   |||||| || ||||  ||||||||  ||
Db    901 CATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACTACAACTGGAGGAAA 960

Qy   1075 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1134
          |||||||  || ||  | || |  || |||||||| |||||||||  | || || ||||||
Db    961 CATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCTTAGCGTGTGACCAT 1020

Qy   1135 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194
          || |||||||||||||||| || || || || ||| |||||||| ||||| ||||| |||
Db   1021 GTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGTGGTTCTGGAACTAA 1080
```

```
RESULT 12
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN
```

```
  Query Match             12.4%;  Score 269;  DB 5;  Length 1567;
  Best Local Similarity   57.8%;  Pred. No. 3.7e-36;
  Matches  523;  Conservative    0;  Mismatches  370;  Indels   12;  Gaps    2;

Qy    297 TAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATC 356
          | | ||||| | | |  || | ||||| |  || || ||       |||||| |   |
Db    461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy    357 CTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAA 416
          ||      |  |||||| ||||| || || ||| | |||   ||| |    || | |
Db    521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy    417 CAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACC 476
          ||| | |||  ||  || ||||||| |||||  | |||  ||||          || ||
Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGGAT---------TTGCC 631

Qy    477 TCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCA 536
          || |   |  | ||| | || || || ||||| |||  |||| | ||||| || | | |
Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy    537 CACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGA 596
          |||  | ||| ||    ||  |  |  ||| || |||||   || ||  |||   ||
Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy    597 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTT 653
          |||||||| | |||||||||    | |  || |    |  | |    | |         ||
Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy    654 TGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 713
          | || |  ||||| ||  | || |||||| | || |||||||   |     | ||
Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy    714 TGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCA 773
          |     |||| || || ||| |    ||    ||| || |  |  |||||||
Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy    774 AGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATT 833
          |||     |    || || ||| |  ||             |||| | |||||||||
Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy    834 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 893
          || ||||| ||||| ||||| ||||||| ||| ||  |||| || ||||| | |||| |
Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051
```

```
Qy    894 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 953
Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111
Qy    954 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAA 1013
Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171
Qy   1014 ATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAA 1073
Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231
Qy   1074 GGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCA 1133
Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291
Qy   1134 TGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1193
Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351
Qy   1194 AAGTG 1198
Db   1352 ATGGG 1356

RESULT 13
AB023628
LOCUS       AB023628                 1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES    Location/Qualifiers
     source 1. .1832
            /organism="Danio rerio"
            /db_xref="taxon:7955"
     gene   332. .1510
            /gene="zFT2"
     CDS    332. .1510
            /gene="zFT2"
            /codon_start=1
            /product="alpha(1,3)fucosyltransferase"
            /protein_id="BAA76707.1"
            /db_xref="GI:4587298"
            /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
            TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
            IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
            QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
            DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
            FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
            LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
            "
BASE COUNT      585 a    394 c    317 g    536 t

ORIGIN

  Query Match          10.6%;  Score 230.4;  DB 5;  Length 1832;
  Best Local Similarity  54.9%;  Pred. No. 1.2e-29;
  Matches  503;  Conservative    0;  Mismatches  401;  Indels   12;  Gaps    2;
Qy    297 TAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATC 356
Db    619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678
Qy    357 CTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAA 416
Db    679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738
Qy    417 CAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACC 476
Db    739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGCACAGAT---------TTTCC 789
Qy    477 TCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCA 536
Db    790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849
Qy    537 CACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGA 596
Db    850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909
Qy    597 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA---AATCCCTTTGTGTT 653
Db    910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969
Qy    654 TGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 713
Db    970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029
Qy    714 TGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCA 773
Db   1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089
Qy    774 AGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATT 833
Db   1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149
Qy    834 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 893
Db   1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209
Qy    894 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 953
Db   1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269
Qy    954 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAA 1013
Db   1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329
Qy   1014 ATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAA 1073
Db   1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389
Qy   1074 GGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCA 1133
Db   1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449
Qy   1134 TGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1193
Db   1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509
Qy   1194 AAGTGTCCATCATTGT 1209
Db   1510 ATGGGATTATGAATAT 1525

RESULT 14
AF288369
```

```
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN
```

Query Match 10.2%; Score 220.6; DB 5; Length 1264;
Best Local Similarity 57.2%; Pred. No. 5.8e-28;
Matches 441; Conservative 0; Mismatches 324; Indels 6; Gaps 2;

```
Qy    307 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 366
Db    175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy    367 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426
Db    235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy    427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 486
Db    295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy    487 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 546
Db    355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy    547 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 606
Db    415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy    607 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 666
Db    475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy    667 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 726
Db    535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy    727 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 786
```

```
Db    595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGCGGCAACAC---TTG 651

Qy    787 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 846
Db    652 CCCCTGTCCATGGACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy    847 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTAC---AATGCATTTTTGGCT 903
Db    712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

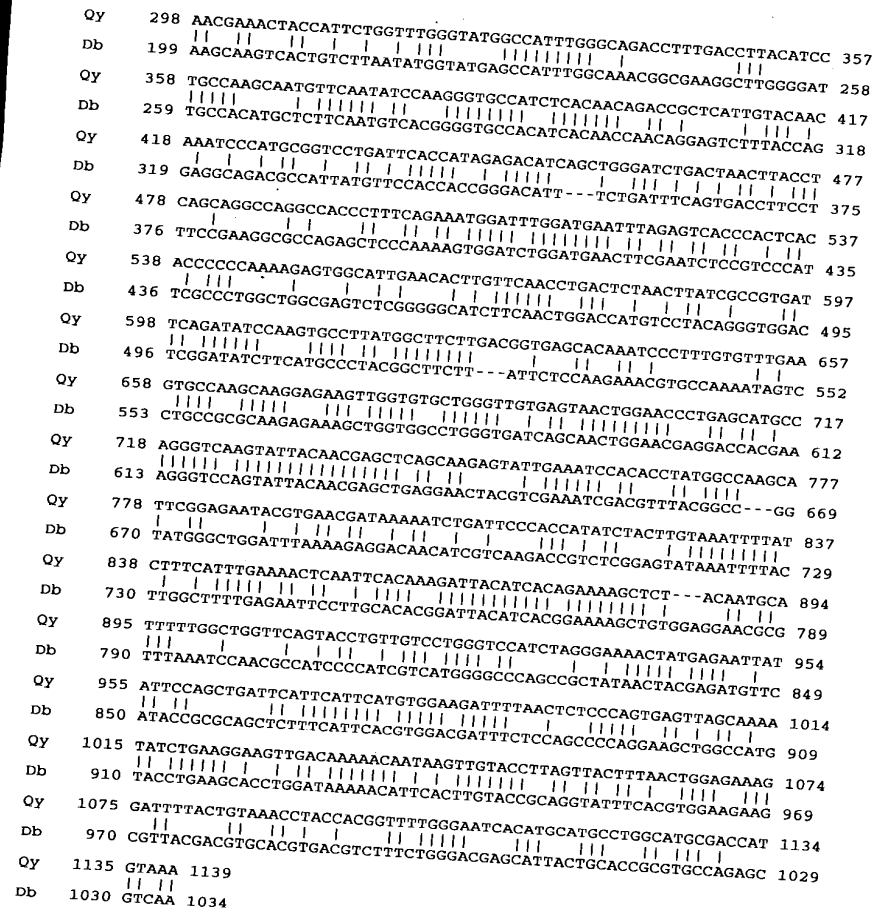
Qy    904 GGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCT 963
Db    772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy    964 GATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAG 1023
Db    832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy   1024 GAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
Db    892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942
```

```
RESULT 15
AB035907
LOCUS       AB035907               1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN
```

Query Match 9.5%; Score 206.6; DB 5; Length 1092;
Best Local Similarity 56.1%; Pred. No. 1.4e-25;
Matches 474; Conservative 0; Mismatches 359; Indels 12; Gaps 4;



Search completed: October 6, 2002, 04:09:47
Job time : 3282.24 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 243.519 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-3
Perfect score: 2170
Sequence: 1 ctttccctcgctgggatcgc.....aaaaaaaaaaaaaaaaaaaa 2170

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2170	100.0	2170	21	AAZ92645	Murine alpha-1,3-f
2	2036	93.8	2036	21	AAZ92672	Murine alpha-1,3-f
3	1206.8	55.6	2676	21	AAZ92647	Human alpha-1,3-fu
4	1092.2	50.3	2822	21	AAZ92646	Human alpha-1,3-fu
5	129	5.9	1814	18	AAT59505	Murine myeloid-lin
6	128.6	5.9	3594	18	AAT85219	Mouse alpha-fucosy
7	128.2	5.9	1086	15	AAQ56912	Lewis blood group
8	128.2	5.9	2042	18	AAT80111	Fucosyltransferase
9	128.2	5.9	2043	12	AAQ13330	Glycosyltransferas
10	128.2	5.9	2043	15	AAQ56906	DNA encoding a gly
11	128.2	5.9	2043	18	AAT76769	Human alpha 1,3/4
12	128.2	5.9	2043	18	AAT61675	Human alpha(1,3/1,
13	126.6	5.8	1316	15	AAQ56911	DNA encoding a gly
14	126.6	5.8	1316	18	AAT61679	Human alpha(1,3)-f
15	124.6	5.7	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
16	124.6	5.7	1654	18	AAT61680	Human alpha(1,3)-f
17	123	5.7	1126	21	AAF21135	Human low adenosin
18	123	5.7	1126	21	AAA35013	Human adenosine re
19	123	5.7	6944	21	AAF21137	Human low adenosin
20	123	5.7	6944	21	AAA35015	Human adenosine re
21	122.8	5.7	1701	15	AAQ77732	Human alpha-1,3-fu
22	122.8	5.7	1701	21	AAF21136	Human low adenosin
23	122.8	5.7	1701	21	AAA35014	Human adenosine re
24	120	5.5	1256	21	AAF21134	Human low adenosin
25	120	5.5	1256	21	AAA35012	Human adenosine re
26	120	5.5	1400	13	AAQ31436	Encodes a HeLa cel
27	120	5.5	1488	15	AAQ56910	DNA encoding a gly
28	120	5.5	2175	12	AAQ14382	Clone 7.2 encoding
29	120	5.5	2175	24	AAS17082	Human cDNA clone 7
30	120	5.5	2861	12	AAQ14383	Clone 1 encoding 1
31	120	5.5	2861	21	AAF21133	Human low adenosin
32	120	5.5	2861	21	AAA35011	Human adenosine re
33	120	5.5	2861	24	AAS17083	Human cDNA clone 1
34	120	5.5	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
35	120	5.5	3647	15	AAQ56909	DNA encoding a gly
36	120	5.5	3647	18	AAT61678	Human alpha(1,3)-f
37	118.4	5.5	2175	11	AAQ06691	CDX, a MILA involv
38	116.8	5.4	2134	18	AAT59506	Human myeloid deri
39	116	5.3	795	17	AAT13798	Alpha-(1,3/1/4)-fu
40	87	4.0	87	21	AAZ92671	Murine alpha-1,3-f
c 41	80.4	3.7	19380	24	AAS61426	Human gene regulat
c 42	78	3.6	15548	24	ABL34155	Human immune syste
c 43	77.4	3.6	15373	24	ABL32467	Human immune syste
c 44	74.6	3.4	9810	24	ABL32427	Human immune syste
c 45	74.2	3.4	7167	24	ABL32400	Human immune syste

ALIGNMENTS

RESULT 1
AAZ92645
ID AAZ92645 standard; cDNA; 2170 BP.
XX
AC AAZ92645;
XX
DT 05-JUN-2000 (first entry)
XX
DE Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; murine; mouse; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 115..1194
FT /*tag= a
FT /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
DR P-PSDB; AAY80995.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Claim 4; Page 127-134; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC (AAY80995), and sequences AAZ92670-Z92673 are murine
CC alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC amplification of cDNA ends).
XX
SQ Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

Query Match 100.0%; Score 2170; DB 21; Length 2170;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60

Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120

Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180

Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240

Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300

Qy     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360

Qy     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420

Qy     421 TCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAG 480

Qy     481 CAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACC 540

Qy     541 CCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCA 600

Qy     601 GATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTG 660

Qy     661 CCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGG 720

Qy     721 GTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTC 780

Qy     781 GGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTT 840

Qy     841 TCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTG 900

Qy     901 GCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCA 960

Qy     961 GCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 1020

Qy    1021 AAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTT 1080

Qy    1081 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1140

Qy    1141 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAGTGTC 1200

Qy    1201 CATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA 1260

Qy    1261 AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATAT 1320

Qy    1321 TTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCC 1380

Qy    1381 TGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAACAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAACAT 1440

Qy    1441 TGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTACTTGTTTCTACATCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTACTTGTTTCTACATCA 1500
```

```
Qy   1501 ATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTT 1560

Qy   1561 TGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTG 1620

Qy   1621 AATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGATAGAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGATAGAG 1680

Qy   1681 TTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATAT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATAT 1740

Qy   1741 GAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGA 1800

Qy   1801 TTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTT 1860

Qy   1861 AGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTT 1920

Qy   1921 TGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCT 1980

Qy   1981 TATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATACCATTTAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATACCATTTAT 2040

Qy   2041 AGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCATCCATATTTAAGTTT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCATCCATATTTAAGTTT 2100

Qy   2101 TATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTGAAAAAAAAAAAAAAAAAAAAA 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTGAAAAAAAAAAAAAAAAAAAAA 2160

Qy   2161 AAAAAAAAAA 2170
          ||||||||||
Db   2161 AAAAAAAAAA 2170
```

RESULT 2
AAZ92672
ID AAZ92672 standard; cDNA; 2036 BP.
XX
AC AAZ92672;
XX
DT 05-JUN-2000 (first entry)
XX
DE Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; murine; mouse; ss.
XX
OS Mus sp.
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer.
XX
PS Example 8; Page 162-164; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC (AAY80995), and sequences AAZ92670-Z92673 are murine
CC alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC amplification of cDNA ends).
XX
SQ Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

Query Match 93.8%; Score 2036; DB 21; Length 2036;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2036; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    107 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 60

Qy    167 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 120

Qy    227 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 180

Qy    287 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 240

Qy    347 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 300

Qy    407 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 360

Qy    467 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 526
```

```
Db    361 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 420

Qy    527 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 586

Db    421 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 480

Qy    587 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 646

Db    481 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 540

Qy    647 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 706

Db    541 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 600

Qy    707 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 766

Db    601 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 660

Qy    767 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 826

Db    661 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 720

Qy    827 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 886

Db    721 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 780

Qy    887 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 946

Db    781 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 840

Qy    947 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 1006

Db    841 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 900

Qy   1007 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 1066

Db    901 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 960

Qy   1067 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1126

Db    961 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1020

Qy   1127 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1186

Db   1021 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1080

Qy   1187 GGAATTAAAGTGTCCATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTT 1246

Db   1081 GGAATTAAAGTGTCCATCATTGTCACAGTGAAAAAGATTAATGAGACATAATTCAAGTTT 1140

Qy   1247 TCAGGATAAGAAGAAACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTG 1306

Db   1141 TCAGGATAAGAAGAAACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTG 1200

Qy   1307 AGGACCATGTATATTTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTG 1366

Db   1201 AGGACCATGTATATTTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTG 1260

Qy   1367 GTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTC 1426

Db   1261 GTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTC 1320

Qy   1427 TCATTTATAAACATTGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTAC 1486

Db   1321 TCATTTATAAACATTGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATGATTTAC 1380

Qy   1487 TTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAA 1546

Db   1381 TTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAA 1440

Qy   1547 TATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACA 1606

Db   1441 TATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACA 1500

Qy   1607 AAGAAGAATTATTGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACT 1666

Db   1501 AAGAAGAATTATTGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACT 1560

Qy   1667 GACAACTGATAGAGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCA 1726

Db   1561 GACAACTGATAGAGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCA 1620

Qy   1727 AATAAGAAAGATATGAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGT 1786

Db   1621 AATAAGAAAGATATGAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGT 1680

Qy   1787 TTAGCAAATGTAGATTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACA 1846

Db   1681 TTAGCAAATGTAGATTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACA 1740

Qy   1847 ATTTATATGGTTTTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATAC 1906

Db   1741 ATTTATATGGTTTTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATAC 1800

Qy   1907 CTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAAT 1966

Db   1801 CTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAAT 1860

Qy   1967 AAATAATTTCCTCTTATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAA 2026

Db   1861 AAATAATTTCCTCTTATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAA 1920

Qy   2027 ATTATACCATTTATAGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCAT 2086

Db   1921 ATTATACCATTTATAGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCAT 1980

Qy   2087 CCATATTTAAGTTTTATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTGAAA 2142

Db   1981 CCATATTTAAGTTTTATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTGAAA 2036
```

```
RESULT 3
AAZ92647
ID   AAZ92647 standard; cDNA; 2676 BP.
XX
AC   AAZ92647;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             295..1374
FT                   /*tag= a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
```

DR WPI; 2000-183120/16.
DR P-PSDB; AAY80996.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Claim 4; Page 143-150; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC AAZ92646 and AAZ92647 represent cDNAs encoding human
CC alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

Query Match 55.6%; Score 1206.8; DB 21; Length 2676;
Best Local Similarity 78.4%; Pred. No. 1.9e-192;
Matches 1674; Conservative 0; Mismatches 387; Indels 75; Gaps 16;

```
Qy       9 CGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGAGTAGTTCA 68
          | || ||| || |||||||||| |||||||| | || |||||  || ||| ||||||  |
Db     187 CCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATA 246

Qy      69 GCACTCATCTC--TGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACA 126
             ||    |||   ||||| ||||| |||||||||| | ||||||||||||||||||||
Db     247 ACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACA 306

Qy     127 TCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCA 186
           |||||||| |||||||||||||| |||| |||||||| |||||||||| |||||||||
Db     307 TCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCA 366

Qy     187 TGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCA 246
           ||||| |||||||| ||||| || |||||||||||| ||||||||||||||||| || ||
Db     367 TGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCC 426

Qy     247 AGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACT 306
           || |||||||||||||||||| ||||| ||||| |||||||||||||||| ||||||
Db     427 AGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACT 486

Qy     307 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 366
           || |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db     487 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 546

Qy     367 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426
           |||||||| |||||||| |||||||||||||| ||||| ||| ||||||||||||||||
Db     547 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606

Qy     427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 486
           || || ||||| || ||  |||||||||| |||||||||| || ||||||||||| ||
Db     607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 666

Qy     487 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 546
           |||||||||| |||||||||||||||||||||||| || ||||| |||||||| ||||||
Db     667 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 726

Qy     547 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 606
           |||||||||||||| |||||||| |||||||||| ||||| ||||||||||||||||||
Db     727 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 786

Qy     607 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 666
           |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db     787 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 846

Qy     667 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 726
           || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     847 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 906

Qy     727 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 786
           ||||||| |||| ||||| || |||||||||| ||||| || |||||||| ||||||
Db     907 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 966

Qy     787 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 846
           || || || |||||||||| ||||||| |||||||| |||||||||||||||||||| |||
Db     967 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 1026

Qy     847 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 906
           ||||| ||||| ||||| |||||||||| |||||||| |||||||| ||| |||||||
Db    1027 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 1086

Qy     907 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 966
           || |||||||||| ||||| |||||||||||||||||||||||||||||||||||| |||
Db    1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146

Qy     967 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 1026
           |||||||||||||||||||| |||||||||||||||| ||||||| |||||||||||||
Db    1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 1206

Qy    1027 GTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTA 1086
           || |||||||||||||||| |||||||||||||||||||||| |||||||| ||||||
Db    1207 GTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTA 1266

Qy    1087 AACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCAT 1146
           || || ||||| |||||||||||||||||||||  |||| ||||| ||||| |||||||||
Db    1267 AATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCAT 1326

Qy    1147 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT-AAAGTGTCCATCA 1205
           |||||||||||||||||||||||||||||||||||||||||||||| ||| | | |||||
Db    1327 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCA 1386

Qy    1206 TTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA-- 1260
             |  ||||   |  ||| |  || |||||| || || ||||| || ||||||||||||
Db    1387 CTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGA 1446

Qy    1261 ---AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTA 1317
              ||||  || | |||| | |||  |||| ||||    ||||| || | | | | ||||
Db    1447 TGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTA 1506

Qy    1318 TATTTTGGT-GAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATT 1376
           ||||||||| || |||||||| || |||| || |||||||| || ||| ||||||||||
Db    1507 TATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATT 1566

Qy    1377 ATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTTATTTGTCACTCTCATTTATA 1435
           |||||||||||  | || | | |||||||| | |||||||||  ||| | |||||| ||
Db    1567 ATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTA 1626

Qy    1436 AACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTAAATTTATGATT---TACTTG 1489
           ||||||| |||| ||   |||| |||||| |  ||| ||||  || || ||   |  |||
Db    1627 AACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTG 1686

Qy    1490 TTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATAT 1549
           |||| |||   ||||| | |||||||||||||||| || || || || ||| |||| |||
Db    1687 TTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTAT 1746
```

```
Qy   1550 GAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAG 1609
          ||   | ||||  | ||||||||| || || ||| | | |  ||| ||| |   |   | |
Db   1747 GA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGG 1805

Qy   1610 AAGAATTATTGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGAC 1669
          |||||    ||   |  || | |  ||| |||||||||| | ||| ||||        |
Db   1806 AAGAA--CATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTGTAT 1863

Qy   1670 AACTG-------------ATAGAGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAA 1715
          |||||             ||    |||     | || ||||||||||||| ||  | |
Db   1864 AACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAATTA 1923

Qy   1716 TTGATGTAGCAAATAAGAAAGATATGAAGCAGAAATGT-------------TATTAGAC 1761
          || ||| ||  |||||||||||||||||||||||| |||            ||||||||
Db   1924 TTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTAGAC 1983

Qy   1762 TTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGATTCCCAATATTTTAAGTTAAA 1821
          ||| |||| ||||| |||||| || || |||  |    |||| ||   | |||| |||||
Db   1984 TTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATTAAA 2043

Qy   1822 GAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTTAGGGAGATTTGATTGTTTGTT 1881
            | |||| ||||| | |||  |||||| |   | ||           ||
Db   2044 TTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATT---------------TTCAGC 2087

Qy   1882 TTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCA 1941
          ||||        || |||||||| | ||| |||| |  || |||||  | |   |
Db   2088 ACCTGGGAAGTAATCCCAATAATACTTTAGAAAATC-TAAGACAGTTCTTTCTGCTACTG 2146

Qy   1942 TGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCTTATAACAATAAAAAAATGATA 2001
            | |       |  |   |||| ||||||||||||  |||||||  |||||
Db   2147 ATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAA------ 2200

Qy   2002 TCTTCAAATACAAATAAATGTATAAATTATACCATTTATAGAACGCTAGTTACT-TCATT 2060
               |  || ||||| || |||||||  |  |||||||  |       | | | |   |
Db   2201 -TGATACCTATAAATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGTAT 2259

Qy   2061 CATTTGAAATATTTTTTCCAAAGCATCCATATTTAA 2096
          || || ||   |||||||  ||||    ||||||||
Db   2260 CAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAA 2295
```

RESULT 4
AAZ92646
ID AAZ92646 standard; cDNA; 2822 BP.
XX
AC AAZ92646;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1289..2368
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
DR P-PSDB; AAY80996.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer.
XX
PS Claim 4; Page 134-142; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC AAZ92646 and AAZ92647 represent cDNAs encoding human
CC alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

Query Match 50.3%; Score 1092.2; DB 21; Length 2822;
Best Local Similarity 85.8%; Pred. No. 2.4e-173;
Matches 1309; Conservative 0; Mismatches 198; Indels 18; Gaps 8;

```
Qy    107 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 166
          |||||||||||||||||||||||||||||| |||||||||||||| |||| ||||||| |
Db   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340

Qy    167 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 226
          |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400

Qy    227 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 286
          |||||||||||||| || || || |||||||||||||||||||| ||||| ||||| ||||
Db   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460

Qy    287 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 346
          |||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||||
Db   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520

Qy    347 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 406
          |||||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |
Db   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580

Qy    407 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 466
          || |||||||||||||||||| || ||||| || ||  |||||||||| ||||||||||
Db   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640

Qy    467 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 526
          | || |||||||||| || |||||||||| |||||||||||||||||||||||||| || |
Db   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700

Qy    527 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 586
```

```
Db   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760

Qy    587 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 646

Db   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820

Qy    647 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 706

Db   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880

Qy    707 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 766

Db   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940

Qy    767 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 826

Db   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000

Qy    827 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 886

Db   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060

Qy    887 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 946

Db   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120

Qy    947 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 1006

Db   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180

Qy   1007 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 1066

Db   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240

Qy   1067 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1126

Db   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300

Qy   1127 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1186

Db   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360

Qy   1187 GGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAA 1242

Db   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420

Qy   1243 GTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGC 1297

Db   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480

Qy   1298 CCTCCCTTGAGGACCATGTATATTTTGGTGAA-ATTTTTAAGAGATCAGAATTAGCAATC 1356

Db   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540

Qy   1357 ACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTT 1415

Db   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600

Qy   1416 ATTTGTCACTCTCATTTATAAACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTA 1472

Db   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660

Qy   1473 AATTTATGATT---TACTTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAA 1529

Db   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720

Qy   1530 AAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAG 1589

Db   2721 AGATGCACATCTTAAAGTATGA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAAC 2779

Qy   1590 CTTGTATAATGCTAACAAAGAAGAA 1614

Db   2780 TTTGCATACTATAACAGAGGAAGAA 2804


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag= a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             5.9%;  Score 129;  DB 18;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 6.9e-13;
  Matches  335;  Conservative    0;  Mismatches  300;  Indels    6;  Gaps    2;

Qy    383 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 442

Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy    443 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 502

Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    503 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 562
```

```
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727
Qy    563 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 622
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787
Qy    623 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 682
Db    788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844
Qy    683 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 742
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904
Qy    743 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 802
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964
Qy    803 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 862
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024
Qy    863 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 919
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084
Qy    920 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 979
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144
Qy    980 TGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 1020
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185

RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key
FT   CDS             Location/Qualifiers
FT                   996..3082
FT                   /*tag= a
FT   exon            /note= "includes introns"
FT                   465..477
FT                   /*tag= b
FT   intron          /number= 1
FT                   478..669
FT                   /*tag= c
FT   exon            /number= 1
FT                   670..1149
FT                   /*tag= d
FT                   /number= 2
FT   intron          /codon_start= 996..998
FT                   1150..1946
FT                   /*tag= e
FT   exon            /number= 2
FT                   1947..1959
FT                   /*tag= f
FT                   /number= 3
FT                   /codon_start= 1947..1949
                     /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag= g
FT   exon            /number= 3
FT                   2067..3552
FT                   /*tag= f
FT                   /number= 4
FT                   /codon_start= 2126..2128
XX                   /note= "alternative ATG start codon"
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;    97WO-US03184.
XX
PR   08-MAR-1996;    96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM, Lowe JB, Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match          5.9%;  Score 128.6;  DB 18;  Length 3594;
  Best Local Similarity  52.5%;  Pred. No. 8.4e-13;
  Matches  331;  Conservative    0;  Mismatches  294;  Indels    6;  Gaps    2;
Qy    383 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 442
Db   2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336
Qy    443 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 502
Db   2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396
Qy    503 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 562
Db   2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456
Qy    563 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 622
Db   2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516
Qy    623 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 682
Db   2517 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573
```

```
Qy    683 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 742
Db   2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy    743 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 802
Db   2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy    803 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 862
Db   2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy    863 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 919
Db   2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy    920 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 979
Db   2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy    980 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 1010
Db   2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904


RESULT 7
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
DE   Lewis blood group fucosyltransferase DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding a Lewis blood group fucosyl
CC   transferase (Fuc-TIII).  The enzyme produced by the DNA may be
CC   non glycosylated.  This prevents premature loss of enzyme activity.
CC   It can also be used in in vitro reactions to modify cell surface
CC   oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-11.
XX
SQ   Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

  Query Match             5.9%;  Score 128.2;  DB 15;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 9.2e-13;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 8
AAT80111
ID   AAT80111 standard; cDNA; 2042 BP.
XX
AC   AAT80111;
XX
DT   01-DEC-1997  (first entry)
XX
DE   Fucosyltransferase III coding sequence.
XX
KW   P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW   calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW   thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW   inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW   disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW   adult respiratory distress syndrome; tumour metastasis;
KW   fucosyltransferase III; acetylglucosaminyltransferase;
KW   autoimmune disorder; inflammatory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9706176-A2.
XX
PD   20-FEB-1997.
XX
```

```
PF   02-AUG-1996;   96WO-US12820.
XX
PR   17-MAY-1996;   96US-0649802.
PR   03-AUG-1995;   95US-0510920.
PR   15-MAY-1996;   96US-0017794.
XX
PA   (OKLA ) UNIV OKLAHOMA STATE.
XX
PI   Cummings RD,  Mcever RP,  Moore KL;
XX
DR   WPI; 1997-154206/14.
XX
PT   New O-glycan(S) derived from P-selectin glyco:protein ligand
PT   (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT   for treating inflammatory or autoimmune disorders or tumours
XX
PS   Disclosure; Page 80-81; 96pp; English.
XX
CC   This sequence represents the coding sequence for the human
CC   fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC   acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC   expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC   mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC   protein expressed on the surfaces of activated platelets and endothelium
CC   in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC   high affinity P-selectin ligand produced by leukocytes. Binding of
CC   P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC   treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC   highly extended extracellular domain, which is a feature of mucin-like
CC   proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC   the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC   O-glycans, and O-glycanated products of the invention. The products are
CC   used for inhibiting the binding of P-selectin and other selectins to
CC   PSGL-1. They can be used for the treatment of inflammation, injury
CC   resulting from ischaemia and reperfusion, bacterial sepsis and
CC   disseminated intravascular coagulation, adult respiratory distress
CC   syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC   other autoimmune or inflammatory disorders. The products can also be used
CC   in the production of antibodies and in diagnostic applications.
XX
SQ   Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

  Query Match                    5.9%;  Score 128.2;  DB 18;  Length 2042;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-13;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | ||||||   |||||    |  |  |  |||||  |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || ||||   |    |   |  | | | | || |     | |||||  |       |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| |  || || ||| |||||  || ||  ||  ||    |  | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  | | | |   |   ||              | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |   ||                |  |  |  ||  || || |   |   |   |  |||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
          | |   |   ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || |||||||| || ||||| |              || ||  |  |||  | |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||      ||||| |||| |  | ||| | || |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    ||||  ||  | ||     || ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |      |||||  || ||||||  ||||  |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044

RESULT 9
AAQ13330
ID   AAQ13330 standard; DNA; 2043 BP.
XX
AC   AAQ13330;
XX
DT   07-NOV-1991  (first entry)
XX
DE   Glycosyltransferase gene in clone pCDM7-alpha(1,3/1,4)FT.
XX
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key
FT   CDS              Location/Qualifiers
FT                    73..1158
FT   polyA_site       /*tag=  a
FT                    2023..2034
FT   polyA_signal     /*tag=  b
FT                    1926..2003
FT   repeat_region    /*tag=  c
FT                    1183..1547
FT                    /*tag=  d
FT   repeat_region    /note= "20 copies of a degenerate 16bp sequence"
FT                    1885..1999
FT                    /*tag=  e
XX                    /note= "similar to human Alu sequences"
PN   WO9112340-A.
XX
PD   22-AUG-1991.
XX
PF   14-FEB-1991;   91WO-US00899.
XX
PR   12-DEC-1990;   90US-0627621.
PR   14-FEB-1990;   90US-0479858.
PR   14-FEB-1990;   90US-0480133.
XX
PA   (UNMI ) UNIV OF MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1991-267151/36.
DR   P-PSDB; AAR13749.
XX
PT   Isolation of gene conveying post-translational characteristic -
PT   e.g. the presence of soluble or membrane bound oligo or
PT   polysaccharide or glycosyltransferase.
XX
PS   Claim 11; Fig 1; 155pp; English.
XX
```

```
CC   The DNA encodes a protein sequence capable of functioning as a
CC   GDP- Fuc:[beta-D-Gal(1,4/1,3)]-D-GlcNAc(/Glc)alpha (1,3/1,4)-
CC   fucosyltransferase. The sequence coded by nucleotides 127 to 1083
CC   represents the functional protein. The DNA sequence can be used to
CC   construct animal cell lines with respect to posttranslational
CC   modification of the oligosaccharides on cell-surface, intracellular
CC   or secreted proteins or lipids by sub-terminal alpha(1,3) and alpha
CC   (1,4) fucose residues that represent the products of this enzyme.
CC   These can be used for the prodn. of diagnostics and therapeutics.
CC   See also AAQ13331-Q13333.
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match             5.9%;  Score 128.2;  DB 12;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-13;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
           ||||| |||| |||||||    |||||    |  |  |  ||||| |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
           | || |||| |   |   | | |   | | || |      | |||||      |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
             ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
             | ||||| || ||   | | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
           |   ||  |  | |  |    |  ||           | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
           || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
           |    ||         | |  |  |  ||   || || |   |  |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
             | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
           ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
           || || || ||||| ||    ||    ||||| |||  |   | ||| | || |  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||| |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 10
AAQ56906
ID   AAQ56906 standard; DNA; 2043 BP.
XX
AC   AAQ56906;
XX
DT   26-JUL-1994  (first entry)

XX
DE   DNA encoding a glycosyltransferase.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             72..1158
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45934.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 1; 249pp; English.
XX
CC   The sequence is that encoding human glycosyl transferase.  The enzyme
CC   produced by the DNA may be non glycosylated.  This prevents premature
CC   loss of enzyme activity.  It can also be used in in vitro reactions to
CC   modify cell surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-12.
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match             5.9%;  Score 128.2;  DB 15;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-13;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
           ||||| |||| | ||||||    |||||    |  |  |  ||||| |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
           | || |||| |   |   | | |   | | || |      | |||||      |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
             ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
             | ||||| || ||   | | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
           |   ||  |  | |  |    |  ||           | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
           || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
           |    ||         | |  |  |  ||   || || |   |  |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759
```

```
Qy     793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 11
AAT76769
ID AAT76769 standard; cDNA; 2043 BP.
XX
AC AAT76769;
XX
DT 15-SEP-1997 (first entry)
XX
DE Human alpha 1,3/4 fucosyltransferase cDNA.
XX
KW GDP-L-fucose:beta-D-N-acetylglucosamide 3/4-alpha-L-fucosyltransferase;
KW human; alpha 1,2 FT; transgene; transgenic mouse; animal model;
KW intestinal adhesion; Helicobacter pylori infection; stomach;
KW small intestine; gut; epithelial cell; surface receptor;
KW carbohydrate antigen; gastritis; peptic ulcer; neoplasia;
KW gastric adenocarcinoma; Lewis antigen; fucosylation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 104..1201
FT /*tag= a
XX
PN US5625124-A.
XX
PD 29-APR-1997.
XX
PF 11-JUL-1994; 94US-0273411.
XX
PR 11-JUL-1994; 94US-0273411.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Falk P, Gordon JI;
XX
DR WPI; 1997-258275/23.
DR P-PSDB; AAW23806.
XX
PT Animal model for Helicobacter pylori infection - comprising
PT transgenic mouse expressing human enzyme promoting intestinal
PT adhesion
XX
PS Example 1; Columns 23-26; 24pp; English.
XX
CC A claimed transgenic mouse expresses, in its intestinal epithelial
CC cells, the enzyme human GDP-L-fucose:beta-D-galactoside 2-alpha-L-
CC fucosyltransferase (also called alpha 1,2 FT) or human GDP-L-fucose:
CC beta-D-N-acetylglucosamide 3/4-alpha-L-fucosyltransferase (also
CC called alpha 1,3/4 FT). The enzyme is expressed under the
CC control of a gut epithelial cell-specific promoter and Helicobacter
CC pylori adheres to the transgenic cells. The transgenic mouse and
CC intestinal epithelial cells from it are useful as models for screening
CC compounds for the ability to inhibit adhesion of H. pylori to gut
CC epithelial cells. The present sequence encodes human alpha 1,3/4 FT
CC and was published by Kukowska-Latallo et al., Genes Dev., 4:
CC 1288-1303 (1990).
XX
SQ Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

Query Match 5.9%; Score 128.2; DB 18; Length 2043;
Best Local Similarity 53.0%; Pred. No. 9.5e-13;
Matches 374; Conservative 0; Mismatches 313; Indels 18; Gaps 4;

```
Qy     385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     622 TTCTTGACGGTGAGCACAAATCCCTTTG----------TGTTTGAAGTGCCAAGCAAGGAG 672
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 12
AAT61675
ID AAT61675 standard; cDNA; 2043 BP.
XX
AC AAT61675;
XX
DT 19-JUN-1997 (first entry)
XX
DE Human alpha(1,3/1,4)-fucosyltransferase (Lewis enzyme) cDNA.
XX
KW Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TIII;
KW glycosylation; oligosaccharide; blood group; ss.
XX

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..1158
FT                   /*tag= a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13638.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 3; Page 268-269; 329pp; English.
XX
CC   A cDNA clone (AAT61675) codes for an enzyme (AAW13638) that functions as
CC   a GDP-Fuc:(beta-D-Gal(1,4/1,3))-D-GlcNAc(/Glc) alpha(1,3/1,4)-
CC   fucosyltransferase (Fuc-TIII) which, when expressed by the cloned
CC   cDNA sequence, functions within mammalian cells to generate de novo
CC   expression of specific SSEA-1 (Lewis x), sialyl-Lewis x, Lewis a and
CC   sialyl-Lewis a structures.  The clone was identified by transfecting
CC   COS-1 cells with genomic DNA from human A431 cells, and isolating a
CC   cloned cDNA that detd. expression of surface-localised SSEA-1
CC   structures.  The cDNA can be used to construct animal cell lines
CC   with specific capabilities with respect to post-translational
CC   modification of the oligosaccharides of expressed proteins or
CC   lipids, to produce recombinant Fuc-TIII for use in oligosaccharide
CC   prodn., to genotype individuals for the Lewis locus, and in the
CC   prodn. of chimeric fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 2043 BP; 413 A; 646 C; 576 G; 408 T; 0 other;

  Query Match             5.9%;  Score 128.2;  DB 18;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-13;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | ||||||    |||||     |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || ||||  |   |   |  | |   | | || |      | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
            ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
            || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |    ||         | |  |  |  ||   || || |   |   |    |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
            | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 13
AAQ56911
ID   AAQ56911 standard; DNA; 1316 BP.
XX
AC   AAQ56911;
XX
DT   26-JUL-1994  (first entry)
XX
DE   DNA encoding a glycosyltransferase.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45939.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding human glycosyl transferase.  The enzyme
CC   produced by the DNA may be non glycosylated.  This prevents premature
CC   loss of enzyme activity.  It can also be used in in vitro reactions to
CC   modify cell surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-12.
XX
SQ   Sequence 1316 BP; 236 A; 456 C; 372 G; 252 T; 0 other;

  Query Match             5.8%;  Score 126.6;  DB 15;  Length 1316;
  Best Local Similarity   52.9%;  Pred. No. 1.7e-12;
  Matches  373;  Conservative    0;  Mismatches  314;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          ||| |  ||||  | |||  |    |||||    |  |  | |||||| |  | |||||
```

```
Db    430 TGCAACATCACTGCCGACTCCAGTGTGTACCCACAGGCAGACGCGGTCATCGTGCACCAC 489
Qy    445 AGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
Db    490 TGGGATATCATGTACAACCCCAGTGCCAACCTCCCGCCCCCCACCAGGCCGCAGGGGCAG 549
Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
Db    550 CGCTGGATCTGGTTCAGCATGGAGTCCCCCAGCAACTGCCGGCACCTGGAAGCCCTGGAC 609
Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
Db    610 GGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 669
Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
Db    670 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 729
Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
Db    730 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCGGCCAGGGTGCGCTACTAC 789
Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
Db    790 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 846
Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
Db    847 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTATCTGGCCTTCGAGAAC 906
Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
Db    907 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 966
Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
Db    967 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCGCCCGACGCC 1026
Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
Db   1027 TTCATCCACGTGGATGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 1086
Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
Db   1087 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1131

RESULT 14
AAT61679
ID   AAT61679 standard; DNA; 1316 BP.
XX
AC   AAT61679;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TV) DNA.
XX
KW   Alpha(1,3)-fucosyltransferase; Lewis enzyme; Fuc-TV;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             121..1245
FT                   /*tag=  a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;    96WO-US13816.
XX
PR   08-SEP-1995;    95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13642.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 5; Page 287-288; 329pp; English.
XX
CC   A DNA clone (AAT61679) codes for an enzyme (AAW13642) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TV) which can utilise neutral type II oligosaccharide
CC   precursors but which cannot efficiently utilise type I
CC   glycococonjugates.  It was identified in a human lambda phage
CC   genomic DNA library using a probe obtd. from Lewis enzyme cDNA (see
CC   also AAT61675).  The Fuc-TV enzyme, when expressed by the cloned DNA
CC   sequence, functions within mammalian cells to generate de novo
CC   expression of specific cell surface glycoconjugate structures that
CC   are recognised by antibodies against SSEA-1, Lewis x or sialyl
CC   Lewis x.
XX
SQ   Sequence 1316 BP; 236 A; 456 C; 372 G; 252 T; 0 other;

Query Match              5.8%;  Score 126.6;  DB 18;  Length 1316;
Best Local Similarity   52.9%;  Pred. No. 1.7e-12;
Matches  373;  Conservative    0;  Mismatches  314;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
Db    430 TGCAACATCACTGCCGACTCCAGTGTGTACCCACAGGCAGACGCGGTCATCGTGCACCAC 489
Qy    445 AGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
Db    490 TGGGATATCATGTACAACCCCAGTGCCAACCTCCCGCCCCCCACCAGGCCGCAGGGGCAG 549
Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
Db    550 CGCTGGATCTGGTTCAGCATGGAGTCCCCCAGCAACTGCCGGCACCTGGAAGCCCTGGAC 609
Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
Db    610 GGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 669
Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
Db    670 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 729
Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
Db    730 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCGGCCAGGGTGCGCTACTAC 789
Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
Db    790 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 846
Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
Db    847 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTATCTGGCCTTCGAGAAC 906
Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
Db    907 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 966
Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
Db    967 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCGCCCGACGCC 1026
Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
```

```
Db   1027 TTCATCCACGTGGATGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 1086

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1087 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1131


RESULT 15
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             5.7%;  Score 124.6;  DB 15;  Length 1654;
  Best Local Similarity   52.1%;  Pred. No. 3.7e-12;
  Matches  408;  Conservative    0;  Mismatches  354;  Indels   21;  Gaps    5;

Qy    310 ATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 369
          || ||| | |||   ||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    370 TTCAAT---ATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426
           |  |   |     |  ||| |  |||| | ||||||    ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAG 483
          ||||||  |  | |||||  ||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    484 GCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCC 543
```

```
          | ||||  |     |||  |||||| ||| | |   | ||||| || |  |||  |
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    544 CAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGAT 603
          ||   ||  | ||| ||    |  ||||| || ||  |  | || ||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    604 ATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTT 654
          |||      ||| || ||||   ||  |  | |  |     |    ||          | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    655 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 714
          |  |  |   ||||    || |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    715 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 774
          ||||||||    || ||| |    ||          | |  |  |  ||   || ||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    775 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 834
            | |    |  | ||  | |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    835 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTAC---AAT 891
          |||||  | || || ||||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    892 GCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAAT 951
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    952 TATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCA 1011
          |   | ||| | ||  | ||||| || ||||| || ||  |    ||||  ||  | ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1012 AAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGA 1071
            || ||| ||||  | |||||  || |       ||||| || ||||||  |||| |
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1072 AAG 1074
            ||
Db   1095 GAG 1097
```

Search completed: October 5, 2002, 23:09:45
Job time : 268.519 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 1816.88 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-3
Perfect score: 2170
Sequence: 1 ctttccctcgctgggatcgc.....aaaaaaaaaaaaaaaaaaaa 2170

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	679.8	31.3	697	9	BB397507	BB397507 BB397507
	2	639	29.4	649	9	AV338075	AV338075 AV338075
	3	635.8	29.3	650	9	AV375511	AV375511 AV375511
	4	580.8	26.8	607	10	BI102760	BI102760 602888338
	5	449.8	20.7	648	9	BB595164	BB595164 BB595164
	6	429	19.8	625	9	BB649967	BB649967 BB649967
	7	418.4	19.3	716	9	AV725071	AV725071 AV725071
	8	416.8	19.2	767	9	BB640911	BB640911 BB640911
c	9	371.2	17.1	579	10	BM021081	BM021081 ie75f03.x
	10	369.8	17.0	628	9	BB651409	BB651409 BB651409
c	11	304.8	14.0	1101	12	CNS059ST	AL327638 Tetraodon
	12	292.8	13.5	316	9	BB130253	BB130253 BB130253
	13	290.6	13.4	455	9	AI849151	AI849151 UI-M-AJ1-
	14	271.4	12.5	670	9	BB626126	BB626126 BB626126
	15	268	12.4	284	9	BB270301	BB270301 BB270301
	16	266.4	12.3	276	9	BB383634	BB383634 BB383634
	17	265	12.2	568	10	BM021317	BM021317 ie75f03.y
	18	259.8	12.0	317	9	BB272545	BB272545 BB272545
c	19	239.2	11.0	501	12	B41073	B41073 HS-1052-B2-
	20	236.2	10.9	315	9	BB250973	BB250973 BB250973
c	21	179.6	8.3	1100	12	CNS0579Z	AL324368 Tetraodon
	22	175.8	8.1	844	12	CNS01V8V	AL168808 Tetraodon
	23	172.2	7.9	199	9	BB398103	BB398103 BB398103
c	24	171.4	7.9	570	10	BM082484	BM082484 fu25e08.x
	25	169.8	7.8	661	9	BB651647	BB651647 BB651647
	26	169	7.8	367	9	BB843869	BB843869 BB843869
c	27	164.8	7.6	887	12	CNS02Q0S	AL208693 Tetraodon
c	28	159.4	7.3	548	12	AQ772584	AQ772584 HS_5568_A
c	29	156.8	7.2	1016	12	CNS0429P	AL271222 Tetraodon
c	30	148.4	6.8	666	10	BJ050300	BJ050300 BJ050300
	31	145.8	6.7	651	10	BJ036855	BJ036855 BJ036855
	32	142	6.5	640	10	BJ073227	BJ073227 BJ073227
c	33	140.4	6.5	711	12	CNS034XF	AL228012 Tetraodon
	34	127	5.9	533	10	BI604066	BI604066 603243886
c	35	126	5.8	674	10	BE605503	BE605503 fl10h04.x
	36	125.4	5.8	322	9	BB498510	BB498510 BB498510
	37	115.6	5.3	751	9	BB260180	BB260180 BB260180
	38	115.6	5.3	971	12	CNS047IV	AL278032 Tetraodon
	39	111.6	5.1	1101	12	CNS00EVL	AL069706 Drosophil
c	40	110.8	5.1	1023	12	CNS02ANG	AL188773 Tetraodon
c	41	110.2	5.1	615	10	BI326811	BI326811 STR00078
	42	106.2	4.9	487	10	BE863396	BE863396 UI-M-BH0-
	43	106.2	4.9	720	10	BF180380	BF180380 601804564
c	44	105.8	4.9	702	10	BJ043578	BJ043578 BJ043578
c	45	97.6	4.5	1101	12	CNS0021J	AL061936 Drosophil

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN
```

```
  Query Match             31.3%;  Score 679.8;  DB 9;  Length 697;
  Best Local Similarity   99.6%;  Pred. No. 2.8e-68;
  Matches  692;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy     518 ATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGA 577
           | ||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db       4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy     578 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 637
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy     638 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy     698 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 757
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy     758 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy     818 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 877
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy     878 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 937
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy     938 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 997
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy     998 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 1057
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy    1058 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 1117
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy    1118 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 1177
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy    1178 AATGGTTTTGGAATTAAAGTGTCCATCATTGTCAC 1212
           ||||||||||||||||||||||||||||||||||
Db     663 AATGGTTTTGGAATTAAAGTGTCCATCATTGTCAC 697
```

```
RESULT 2
AV338075
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
```

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Location/Qualifiers
source 1. .649
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6430407H24"
/clone_lib="RIKEN full-length enriched, adult male olfactory bulb"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 209 a 105 c 117 g 218 t
ORIGIN

Query Match 29.4%; Score 639; DB 9; Length 649;
Best Local Similarity 99.2%; Pred. No. 1.2e-63;
Matches 642; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy    819 ATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGA 878
          ||| ||||||||  |||  ||||||||||||||||||||||||  ||||||||||||||
Db      2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61

Qy    879 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy    939 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy    999 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 1058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy   1059 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 1118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy   1119 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 1178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy   1179 ATGGTTTTGGAATTAAAGTGTCCATCATTGTCACAGTGAAAAAGATTAATGAGACATAAT 1238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 ATGGTTTTGGAATTAAAGTGTCCATCATTGTCACAGTGAAAAAGATTAATGAGACATAAT 421

Qy   1239 TCAAGTTTTCAGGATAAGAAGAAACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCC 1298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 TCAAGTTTTCAGGATAAGAAGAAACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCC 481

Qy   1299 CTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCAC 1358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 CTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAGAGATCAGAATTAGCAATCAC 541

Qy   1359 TCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGAAATAATTTATT 1418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 TCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGAAATAATTTATT 601

Qy   1419 TGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAGTTGACTG 1465
          |||||||||||||||||||||||||||||||||||||||||||||
Db    602 TGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAGTTGACTG 648
```

RESULT 3
AV375511
LOCUS AV375511 650 bp mRNA linear EST 24-OCT-2001
DEFINITION AV375511 RIKEN full-length enriched, adult male cecum Mus musculus cDNA clone 9130026C15 3' similar to AB015426 Mus musculus mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION AV375511
VERSION AV375511.2 GI:16397991
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 650)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A. , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Nov 14, 1999 this sequence version replaced gi:6423158.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES
source Location/Qualifiers
1. .650
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9130026C15"
/clone_lib="RIKEN full-length enriched, adult male cecum"
/sex="male"
/tissue_type="cecum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 240 a 80 c 88 g 242 t
ORIGIN

Query Match 29.3%; Score 635.8; DB 9; Length 650;
Best Local Similarity 99.5%; Pred. No. 2.7e-63;
Matches 648; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy   1379 CCTGTATATATGTGATAACGAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAAC 1438
          ||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||||||||
Db      1 CCTGTATATATGTAATAACAAGCACTGGAAATAATTTATTTGTCACTCTCATTTATAAAC 1438

Qy   1439 ATTGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATTTGTCACTCTCATTTATAAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ATTGTTTTTACATTTTATAGTTGACTGTAAAGTAAATTTATTTGTCACTCTCATTTATAAAC 120

Qy   1499 CAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTT 1558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTT 180

Qy   1559 TTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTAT 1618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTAT 240

Qy   1619 TGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGATAG 1678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTGATAG 300

Qy   1679 AGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGAT 1738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGAT 360

Qy   1739 ATGAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTA 1798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGAAGCAGAAATGTTATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTA 1798

Qy   1799 GATTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTT 1858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GATTCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTT 1858

Qy   1859 TTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCT 1918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCT 1918

Qy   1919 TTTGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTTGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCT 1978

Qy   1979 CTTATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATT 2029
          ||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     601 CTTATAACAAT-AAAAAATGATATCTTCAAATACAAATAAATGTATAAATT 650
```

RESULT 4
BI102760
LOCUS BI102760 607 bp mRNA linear EST 26-JUN-2001
DEFINITION 602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609 5', mRNA sequence.
ACCESSION BI102760
VERSION BI102760.1 GI:14553653
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 607)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11119 row: 1 column: 10
High quality sequence stop: 606.

FEATURES
source Location/Qualifiers
1. .607
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5043609"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT 179 a 149 c 121 g 157 t 1 others
ORIGIN

Query Match 26.8%; Score 580.8; DB 10; Length 607;
Best Local Similarity 99.2%; Pred. No. 4.5e-57;
Matches 604; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy    337 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 396
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60
```

```
Qy     397 ACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 456
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy     457 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATG 516
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy     517 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 576
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy     577 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy     637 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 696
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy     697 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 756
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy     757 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy     817 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy     877 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGG 936
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db     539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy     937 GAAAACTAT 945
           |||||||||
Db     599 GAAAACTAT 607
```

RESULT 5
BB595164
LOCUS BB595164 648 bp mRNA linear EST 26-OCT-2001
DEFINITION BB595164 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION BB595164
VERSION BB595164.2 GI:16450167
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 648)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Nov 30, 2000 this sequence version replaced gi:11491766.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Location/Qualifiers
source 1. .648
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030004N01"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 141 a 206 c 138 g 163 t
ORIGIN

Query Match 20.7%; Score 449.8; DB 9; Length 648;
Best Local Similarity 99.6%; Pred. No. 2.7e-42;
Matches 451; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 255

Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 315
```

```
Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     316 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 375

Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 435

Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 495

Qy     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     496 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 555

Qy     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     556 CAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 615

Qy     421 TCCCATGCGGTCCTGATTCACCATAGAGACATC 453
           |||||||||||||||||||||||| |||||||
Db     616 TCCCATGCGGTCCTGATTCACCATATAGACATC 648
```

RESULT 6
BB649967

```
LOCUS       BB649967                625 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB649967 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION   BB649967
VERSION     BB649967.1  GI:16484222
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C130099N17"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="16 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT      136 a    198 c    135 g    156 t
ORIGIN
```

```
  Query Match             19.8%;  Score 429;  DB 9;  Length 625;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-40;
  Matches  429;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 256

Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 316

Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 376

Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 436

Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 496

Qy     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 556

Qy     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 616

Qy     421 TCCCATGCG 429
```

```
          |||||||||
Db    617 TCCCATGCG 625


RESULT 7
AV725071
LOCUS       AV725071                716 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION   AV725071
VERSION     AV725071.1  GI:10830091
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .716
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCBLB02"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      219 a    113 c    133 g    249 t      2 others
ORIGIN

  Query Match           19.3%;  Score 418.4;  DB 9;  Length 716;
  Best Local Similarity   84.8%;  Pred. No. 9.2e-39;
  Matches  517;  Conservative    0;  Mismatches   83;  Indels   10;  Gaps    4;

Qy    773 AAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAAT 832
          ||||||| |||||||| || || ||||||||| ||||||| ||||||||| |||||||||
Db      1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy    833 TTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATG 892
          |||||||||| |||||||| ||||| ||||| |||||||||| |||||||| |||||||
Db     61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy    893 CATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATT 952
          | ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||||||||
Db    121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy    953 ATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAA 1012
          |||||||||| |||||||||||||||||||||||| ||||||||||||||| ||||||
Db    181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy   1013 AATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAA 1072
          | |||||||||||||| |||||||||||||||| |||||||||||||||||||||||| |
Db    241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy   1073 AGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACC 1132
          ||||||| |||||||| || ||||| |||||||||||||||||||  |||| ||||| |
Db    301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy   1133 ATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1192
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy   1193 -AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAAGTTTTC 1248
           ||| | | ||||| |  ||||   |  ||| |  || |||||| || || ||||| ||
Db    421 AAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTG 480

Qy   1249 AGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCC 1303
          ||||||||||||     |||| || | |||| | |||  |||| ||||    ||||| |
Db    481 AGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTC 540

Qy   1304 TTGAGGACCATGTATATTTTGGT-GAAATTTTTAAGAGATCAGAATTAGCAATCACTCAA 1362
          | | | | |||||||| ||||| || |||||||| || |||| || |||||||| || |
Db    541 TAGGGTAACGTGTATATCTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCA 600

Qy   1363 TTTGGTTTTA 1372
          ||    ||||
Db    601 TTNNCGTTTA 610


RESULT 8
BB640911
LOCUS       BB640911                767 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB640911 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730093G08 5', mRNA sequence.
ACCESSION   BB640911
VERSION     BB640911.1  GI:16476069
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
```

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES e mouse tissues.
source Location/Qualifiers
1. .767
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730093G08"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 175 a 239 c 176 g 172 t 5 others
ORIGIN

Query Match 19.2%; Score 416.8; DB 9; Length 767;
Best Local Similarity 97.0%; Pred. No. 1.3e-38;
Matches 424; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
Db     233 CTTTCCCTCGCTGGGATCGCTTTANAATGTGATAACTCACGCATCTGATAGCACCGTGGA 292
Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
Db     293 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 352
Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
Db     353 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 412
Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
Db     413 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 472
Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
Db     473 TCTGCAAGTTCTGGGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 532
Qy     301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
Db     533 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCGGC 592
Qy     361 CAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAA 420
Db     593 CAAGCAATGTTCAATATCCAGGGGGGCCATCTTACAACAAAACGGTCATTGGACAACAAA 652
Qy     421 TCCCATGCGGTCCTGAT 437
Db     653 TCCCAGGGGGGCCTGAT 669
```

RESULT 9
BM021081/c
LOCUS BM021081 579 bp mRNA linear EST 30-OCT-2001

DEFINITION ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;, mRNA sequence.
ACCESSION BM021081
VERSION BM021081.1 GI:16535437
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 579)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T , Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
High quality sequence stop: 436.
FEATURES
source Location/Qualifiers
1. .579
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."
BASE COUNT 203 a 103 c 93 g 180 t
ORIGIN

Query Match 17.1%; Score 371.2; DB 10; Length 579;
Best Local Similarity 83.7%; Pred. No. 2.2e-33;
Matches 482; Conservative 0; Mismatches 83; Indels 11; Gaps 5;

```
Qy     886 TACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTAT 945
Db     579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520
Qy     946 GAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAG 1005
Db     519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460
Qy    1006 TTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAAC 1065
Db     459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400
```

```
Qy   1066 TGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCA 1125
          ||||| |||||||| |||||||| || ||||  ||||||||||||||||||  ||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy   1126 TGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 1185
          ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy   1186 TGGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCA 1241
          ||||||| ||| | | ||||| |  ||||   |  ||| |  || |||||| || || ||
Db    279 TGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCA 220

Qy   1242 AGTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTG 1296
          ||| || |  |||||||||     ||||  || | |||| | |||  |||| ||||
Db    219 AGTATTGAAAATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCA 160

Qy   1297 CCCTCCCTTGAGGACCATGTATATTTTGGT-GAAATTTTTAAGAGATCAGAATTAGCAAT 1355
          ||||| || | | | |||||||||||||| || |||||||| || |||| || ||||||
Db    159 CCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAAT 100

Qy   1356 CACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA-AATAATT 1414
          || || ||| |||||||||||||||||||||  | || | | ||||||||  | |||||
Db     99 CATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATT 40

Qy   1415 TATTTGTCACTCTCATTTATAAACATTGTTTTTACA 1450
          ||||  ||| | |||||| |||||||| |||| ||
Db     39 TATTCTTCATTATCATTTGTAAACATTGCTTTTTCA 4
```

RESULT 10
BB651409
LOCUS BB651409 628 bp mRNA linear EST 26-OCT-2001
DEFINITION BB651409 RIKEN full-length enriched, 0 day neonate cerebellum Mus musculus cDNA clone C230079N02 5', mRNA sequence.
ACCESSION BB651409
VERSION BB651409.1 GI:16485662
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 628)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .628
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230079N02"
/clone_lib="RIKEN full-length enriched, 0 day neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence {5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 129 a 200 c 143 g 152 t 4 others
ORIGIN

Query Match 17.0%; Score 369.8; DB 9; Length 628;
Best Local Similarity 97.4%; Pred. No. 3e-33;
Matches 374; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 292

Qy     61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 352

Qy    121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 412

Qy    181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 472

Qy    241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATGTTAAC 532

Qy    301 GAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 360
          |||||||||||||||| || |||||||||||| |||||||||||| ||||||||||||||||
Db    533 GAAACTACCATTCTGGNTTAGGTATGGCCATNTGGGCAGACCTNTGACCTTACATCCTGC 592

Qy    361 CAAGCAATGTTCAATATCCAAGGG 384
          |||   |||||  ||||||||||||
```

```
Db    593 CAAAGCATGTTTCATATCCAAGGG 616


RESULT 11
CNS059ST/c
LOCUS       CNS059ST                1101 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            045P22 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL327638
VERSION     AL327638.1  GI:8221227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="045P22"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA045DH11A1~end : T3"
BASE COUNT      287 a    249 c    269 g    278 t     18 others
ORIGIN

  Query Match           14.0%;  Score 304.8;  DB 12;  Length 1101;
  Best Local Similarity  58.7%;  Pred. No. 5e-26;
  Matches  608;  Conservative    0;  Mismatches  418;  Indels    9;  Gaps    5;

Qy     164 TCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGG 223
           ||||| |||| |||| || | |  | || ||   || ||  ||| ||||| |
Db    1064 TCATCGTGGGGTGCTNTGTGACTGTGTNNTTTATGTACTNTAATCCATCCACCAGCT-GN 1006

Qy     224 TCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAA 283
           | | | |||| ||||| | | |  | | | || |||   | || || |
Db    1005 TTTCCGGTCCTGTGGAGGCAGGACTGAGCTCGGAGCATGTCAAGAATCAACTATCAACCA 946

Qy     284 AAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCT 343
           | | |||| |   |||  || | || ||| | ||| | ||||| || || || ||||
Db     945 AGAGTGATAAA---AACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCT 889

Qy     344 TTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACC 403
           | ||||| |     ||| | ||   |||   | | || ||  |  | ||  | |||
Db     888 TCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACA 829

Qy     404 GCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATC 463
           |     | |  ||||| ||  |||  ||  |  | ||||| |||||||||  | | || |
Db     828 GAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACC 769

Qy     464 TGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAG 523
           || | ||| | ||        |   | ||  || | ||||||||||| ||| |  |  | |
Db     768 TGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGG 709

Qy     524 AGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAA 583
           |||| ||| | ||| || |      |   ||| || ||   | ||||||||||||| || |
Db     708 AGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGA 649

Qy     584 CTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATC 643
             ||||| |  ||||| || ||         || |  || |     ||  |||  ||
Db     648 ACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATG 589

Qy     644 CCTTTGTG---TTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACT 700
           |   | |   ||||    || |||||  | ||  |||| || ||| |||| || ||||
Db     588 CTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACT 529

Qy     701 GGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCC 760
           ||||||  || |||||  | || || |  ||||||||  |   ||||  | ||||  |||
Db     528 GGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCC 469

Qy     761 ACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATAT 820
           |||| ||||| | |||||||||  || || |   ||| || |     | || |||||
Db     468 ACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCG 409

Qy     821 CTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAA 880
           | |  || || || || || |||||||| ||||| || |||||||| ||| | || ||||
Db     408 CCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAA 349

Qy     881 AGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAA 940
           | || |||||  || ||   |  || |||| || || || ||||| ||| || |||| |
Db     348 AACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGA 289

Qy     941 ACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCA 1000
           |||| |||||    || |  |  ||| | || || || |||||||| |||  |||  |||
Db     288 ACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCA 229

Qy    1001 GTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACT 1060
           |||  | ||  | ||| |   |  | ||||||||  |  |  |||||||||  || |
Db     228 AGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATT 169

Qy    1061 TTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCC 1120
           || |   |  |||   ||   ||   |  ||  ||| |                | || |
Db     168 TTCANGTGGCGAAGACACTTTAAAGCAA-GGGCAGTGGCATTTGCCCGAGGCCACCTGTC 110

Qy    1121 TGGCATGCGACCATGTAAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAA 1179
           |||  || ||  | || |  |||| | |  |||| | ||| |   |  |  ||| | ||
Db     109 AGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAG 50

Qy    1180 TGGTTTTGGAATTAA 1194
           ||||  |||  || |
Db      49 TGGTACTGGGGTTGA 35


RESULT 12
BB130253
LOCUS       BB130253                316 bp    mRNA    linear   EST 28-JUN-2000
DEFINITION  BB130253 RIKEN full-length enriched, 16 days neonate cerebellum Mus
            musculus cDNA clone 9630044I03 3' similar to AB015426 Mus musculus
            Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   BB130253
VERSION     BB130253.1  GI:8784643
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
```

```
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9630044I03"
                     /clone_lib="RIKEN full-length enriched, 16 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="16 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 370.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      115 a     45 c     37 g    119 t
ORIGIN

  Query Match             13.5%;  Score 292.8;  DB 9;  Length 316;
  Best Local Similarity   96.2%;  Pred. No. 2.2e-24;
  Matches  300;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy   1827 CAAGAATTTTCAGATAAACAATTTATATGGTTTTAGGGAGATTTGATTGTTTGTTTTGTG 1886
           | |||||||||||||| ||| ||| ||||||||  |||||||||||||| |||||||| |||
Db      5 CCAGAATTTTCAGATTAACCATTGATATGGTTAGAGGGAGATTTGATCGTTTGTTTCGTG 64

Qy   1887 GGTTTAGTCCCTAATAATACCTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCATGAAG 1946
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||

Db     65 GGTTTAGTCCCTAATAATACCTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCATGGAG 124

Qy   1947 ATAACCACTATTTTAAAAATAAATAATTTCCTCTTATAACAATAAAAAAATGATATCTTC 2006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 ATAACCACTATTTTAAAAATAAATAATTTCCTCTTATAACAATAAAAAAATGATATCTTC 184

Qy   2007 AAATACAAATAAATGTATAAATTATACCATTTATAGAACGCTAGTTACTTCATTCATTTG 2066
           |||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db    185 AAATACAAATAAATGTATAAATTATACCATTCATAGAACGCCAGTTACTTCATTCATTTG 244

Qy   2067 AAATATTTTTTCCAAAGCATCCATATTTAAGTTTTATTTTACTGTATTTGTTAATAAATA 2126
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    245 AAATATTTTTTCCAAAGCATCCATATTTAAGTTTTATTTTACCGTATTTGTTAATAAATA 304

Qy   2127 GAAGTAAAACTT 2138
           ||||||||||||
Db    305 GAAGTAAAACTT 316


RESULT 13
AI849151
LOCUS       AI849151                455 bp    mRNA    linear   EST 15-JUL-1999
DEFINITION  UI-M-AJ1-agy-d-03-0-UI.s1 NIH_BMAP_MOB_N Mus musculus cDNA clone
            UI-M-AJ1-agy-d-03-0-UI 3', mRNA sequence.
ACCESSION   AI849151
VERSION     AI849151.1  GI:5493057
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            The following repetitive elements were found in this cDNA sequence:
            111-188, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AJ1-agy-d-03-0-UI"
                     /clone_lib="NIH_BMAP_MOB_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MOB_N library is a normalized library constructed
                     from  mouse olfactory bulbs. The tag is a string of 5
                     nucleotides present  between the Not I site and the
                     oligo-dT track.  The library was  constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996. Tissue provided by Ms. Annie Novakovich,
                     Zivic-Miller Laboratories.
                     TAG_LIB=NIH_BMAP_MOB_N
```

```
                    TAG_TISSUE=olfactory-bulbs
                    TAG_SEQ=CAAGG"
BASE COUNT      174 a      60 c      60 g     161 t
ORIGIN

  Query Match             13.4%;  Score 290.6;  DB 9;  Length 455;
  Best Local Similarity   95.5%;  Pred. No. 3.2e-24;
  Matches  299;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy   1853 ATGGTTTTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGA 1912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 ATGGTTTTAGGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGA 68

Qy   1913 GAATCTTTTGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAA 1972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 GAATCTTTTGATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAA 128

Qy   1973 TTTCCTCTTATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATA 2032
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 TTTCCTCTTATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATA 188

Qy   2033 CCATTTATAGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCATCCATAT 2092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 CCATTTATAGAACGCTAGTTACTTCATTCATTTGAAATATTTTTTCCAAAGCATCCATAT 248

Qy   2093 TTAAGTTTTATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTGAAAAAAAAAAAAA 2152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |    |
Db    249 TTAAGTTTTATTTTACTGTATTTGTTAATAAATAGAAGTAAAACTTGAAAGATCTATCTG 308

Qy   2153 AAAAAAAAAAAA 2165
           |  |   |||| |
Db    309 ATCATTTAAAAGA 321


RESULT 14
BB626126
LOCUS       BB626126                 670 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB626126 RIKEN full-length enriched, adult male diencephalon Mus
            musculus cDNA clone 9330161D06 5', mRNA sequence.
ACCESSION   BB626126
VERSION     BB626126.1  GI:16464204
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .670
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9330161D06"
                     /clone_lib="RIKEN full-length enriched, adult male
                     diencephalon"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      107 a    246 c    147 g    169 t      1 others
ORIGIN

  Query Match             12.5%;  Score 271.4;  DB 9;  Length 670;
  Best Local Similarity   99.3%;  Pred. No. 3.8e-22;
  Matches  272;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     397 CTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGA 456

Qy      61 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 GTAGTTCAGCACTCATCTCTGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACA 516

Qy     121 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     517 TCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTC 576

Qy     181 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     577 ATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAG 636

Qy     241 TCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCT 274
           ||||  |||||||||||||||||||||||||||
Db     637 TCTGNCAGTTCTGTGCTGAAAATGAAAAATTTCT 670
```

RESULT 15
BB270301
LOCUS BB270301 284 bp mRNA linear EST 07-JUL-2000
DEFINITION BB270301 RIKEN full-length enriched, 10 days neonate cortex Mus musculus cDNA clone A830047F14 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB270301
VERSION BB270301.1 GI:8966773
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 284)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .284
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A830047F14"
/clone_lib="RIKEN full-length enriched, 10 days neonate cortex"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 86 a 42 c 48 g 108 t
ORIGIN

Query Match 12.4%; Score 268; DB 9; Length 284;
Best Local Similarity 96.5%; Pred. No. 1.4e-21;
Matches 274; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy   1171 TTAGAGAAATGGTTTTGGAATTAAAGTGTCCATCATTGTCACAGTGAAAAAGATTAATGA 1230
          |||||| |||||||||||||||||||||||||||| |||||||||| ||||||| ||||
Db      1 TTAGAGTAATGGTTTTGGAATTAAAGTGTCCATCATAGTCACAGTGATAAAGATTTATGA 60

Qy   1231 GACATAATTCAAGTTTTCAGGATAAGAAGAAACACGCTGCATTTGGGGCACCGTTTAATT 1290
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     61 GACATAATTCAAGTTTTCAGGATAAGAAGAGACACGCTGCATTTGGGGCACCGTTTAATT 120

Qy   1291 TTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAGAGATCAGAATTA 1350
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTCCTTCCCTCCCTTGAGGACCATGTATATTTTGGTGAAATTTTTAAGAGATCAGAATTA 180

Qy   1351 GCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGAAAT 1410
          |||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db    181 TCAATCACTCAATTTGGTTTTTAATTATCCTGTATATATGTGATTACGAGCACTGGAAAT 240

Qy   1411 AATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTTT 1454
          ||||||||||||||||||||||||||||||||||||||||||| |
Db    241 AATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTCT 284
```

Search completed: October 6, 2002, 07:15:15
Job time : 1831.88 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 44.2762 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-3
Perfect score: 2170
Sequence: 1 ctttccctcgctgggatcgc.....aaaaaaaaaaaaaaaaaaaa 2170

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128.6	5.9	1814	2	US-08-483-151-1	Sequence 1, Appli
2	128.6	5.9	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	128.2	5.9	1086	1	US-07-914-281-12	Sequence 12, Appl
4	128.2	5.9	1086	1	US-08-393-246-12	Sequence 12, Appl
5	128.2	5.9	1086	1	US-08-525-058A-12	Sequence 12, Appl
6	128.2	5.9	1086	2	US-08-696-731-12	Sequence 12, Appl
7	128.2	5.9	1086	4	US-09-042-531-12	Sequence 12, Appl
8	128.2	5.9	2042	3	US-09-063-237-2	Sequence 2, Appli
9	128.2	5.9	2043	1	US-07-914-281-1	Sequence 1, Appli
10	128.2	5.9	2043	1	US-08-393-246-1	Sequence 1, Appli
11	128.2	5.9	2043	1	US-08-273-411-4	Sequence 4, Appli
12	128.2	5.9	2043	1	US-08-525-058A-1	Sequence 1, Appli
13	128.2	5.9	2043	2	US-08-696-731-1	Sequence 1, Appli
14	128.2	5.9	2043	4	US-09-042-531-1	Sequence 1, Appli
15	128.2	5.9	2043	5	PCT-US91-00899-1	Sequence 1, Appli
16	126.6	5.8	1316	1	US-07-914-281-10	Sequence 10, Appl
17	126.6	5.8	1316	1	US-08-393-246-10	Sequence 10, Appl
18	126.6	5.8	1316	1	US-08-525-058A-10	Sequence 10, Appl
19	126.6	5.8	1316	2	US-08-696-731-10	Sequence 10, Appl
20	126.6	5.8	1316	4	US-09-042-531-10	Sequence 10, Appl
21	124.6	5.7	1654	1	US-07-914-281-13	Sequence 13, Appl
22	124.6	5.7	1654	1	US-08-393-246-13	Sequence 13, Appl
23	124.6	5.7	1654	1	US-08-525-058A-13	Sequence 13, Appl
24	124.6	5.7	1654	2	US-08-696-731-13	Sequence 13, Appl
25	124.6	5.7	1654	4	US-09-042-531-13	Sequence 13, Appl
26	120	5.5	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	120	5.5	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	120	5.5	1488	1	US-07-914-281-9	Sequence 9, Appli
29	120	5.5	1488	1	US-08-393-246-9	Sequence 9, Appli
30	120	5.5	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	120	5.5	1488	2	US-08-696-731-9	Sequence 9, Appli
32	120	5.5	1488	4	US-09-042-531-9	Sequence 9, Appli
33	120	5.5	2175	4	US-08-482-073-9	Sequence 9, Appli
34	120	5.5	2861	4	US-08-482-073-10	Sequence 10, Appl
35	120	5.5	3647	1	US-07-914-281-7	Sequence 7, Appli
36	120	5.5	3647	1	US-08-393-246-7	Sequence 7, Appli
37	120	5.5	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	120	5.5	3647	2	US-08-696-731-7	Sequence 7, Appli
39	120	5.5	3647	4	US-09-042-531-7	Sequence 7, Appli
40	120	5.5	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	116.8	5.4	2134	2	US-08-483-151-3	Sequence 3, Appli
42	116.8	5.4	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	67.4	3.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
44	66.2	3.1	6124	4	US-08-213-419B-3	Sequence 3, Appli
c 45	65.2	3.0	5852	1	US-07-867-106-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 5.9%; Score 128.6; DB 2; Length 1814;
Best Local Similarity 52.5%; Pred. No. 7.8e-17;
Matches 331; Conservative 0; Mismatches 294; Indels 6; Gaps 2;

```
Qy    383 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 442
          | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy    443 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 502
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    503 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 562
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy    563 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 622
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy    623 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 682
Db    788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy    683 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 742
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy    743 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 802
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy    803 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 862
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    863 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 919
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    920 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 979
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    980 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 1010
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06427
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US96-06427-1

Query Match 5.9%; Score 128.6; DB 5; Length 1814;
Best Local Similarity 52.5%; Pred. No. 7.8e-17;
Matches 331; Conservative 0; Mismatches 294; Indels 6; Gaps 2;

```
Qy    383 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 442
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy    443 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 502
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    503 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 562
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy    563 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 622
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy    623 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 682
Db    788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy    683 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 742
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy    743 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 802
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy    803 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 862
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    863 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 919
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    920 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 979
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    980 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 1010
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

RESULT 3
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             5.9%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 8.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          ||||| |||| | |||||| |||||  | | | ||||| | | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || |||| |  |  | | | | | |      | ||||| |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
            ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |   ||  |  | |   |    ||          | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
            || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
            | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 4
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             5.9%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 8.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          ||||| |||| | |||||| |||||  | | | ||||| | | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || |||| |  |  | | | | | |      | ||||| |     |||
```

```
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          ||||| ||| | || || ||| ||||| || ||  ||  ||    |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
            | ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  |  | | |    |   ||          | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || |||||  |||||  |||  ||||||||| || ||    ||||||||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
            |    ||          | |  |  ||   || || |    |    |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCCTCCCACA---AGCCCCTG 687

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
            | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    |||| |||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    |||| ||  | ||    || ||| |||| |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 5
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             5.9%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 8.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | ||||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
            | || ||||  |   |   |  | |  | | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || ||| ||||| || ||  ||  ||    |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
            | ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  |  | | |    |   ||          | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || |||||  ||||  ||| ||||||| || ||    |||||||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
            |    ||          | |  |  ||   || || |    |    |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCCTCCCACA---AGCCCCTG 687

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
            | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    |||| |||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    |||| ||  | ||    || ||| |||| |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 6
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
```

```
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             5.9%;  Score 128.2;  DB 2;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 8.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | ||||||    |||||     |  |  |  |||||  |  | ||||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
           | || ||||  |   |   |  | |   | | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
            ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
            |  ||||| || ||  |  | || ||| | || || || ||| |||  ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |   ||  |  | |  |    |   ||          | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
            || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
          | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    |||| ||  | ||    || ||| |||| |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 7
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match                    5.9%;  Score 128.2;  DB 4;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 8.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          ||||||  ||||  | ||||||     |||||     |  |  |  |||||  |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || ||||  |  |   |  | | | || |  |  |  |||||  |     |||  || |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || ||| ||||| || ||  ||  |  |  | | | | ||| |||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  ||||| || ||  |  || ||| | || || ||| ||  |  |  |  |  | ||||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  |  | |  |  |||  |||         |  |  |  |  |  |  ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || ||||  |||||  |||  |||||||| || ||    |||||||  |  |  |||  |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |  ||  |  | | |  ||  || || |  |  |  |  |  |  |  |  |  |  |||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
          |  |   |  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || |||||||| || |||||  |     |  |  |  |  |  |||  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || |||||  ||   ||  |||||  |||  |  |  |  |  |  |  |  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          |||||  |  |||||  || ||  |  ||||  ||  |  ||  |  |  |  |  |  |||
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  |  |  |  |||||  |||||  ||||  |  |  |||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 8
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
  GENERAL INFORMATION:
    APPLICANT:  McEver, Rodger P.
    APPLICANT:  Cummings, Richard D.
    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
    NUMBER OF SEQUENCES:  5
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Patrea L. Pabst
      STREET:  2800 One Atlantic Center, 1201 West Peachtree
      STREET:  Street
      CITY:  Atlanta
      STATE:  Georgia
      COUNTRY:  US
      ZIP:  30306-3450
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/09/063,237
      FILING DATE:
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/649,802
      FILING DATE:
    ATTORNEY/AGENT INFORMATION:
      NAME:  Pabst, Patrea L.
      REGISTRATION NUMBER:  31,284
      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (404)873-8794
      TELEFAX:  (404)873-8795
  INFORMATION FOR SEQ ID NO:  2:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  2042 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  double
      TOPOLOGY:  linear
    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match                    5.9%;  Score 128.2;  DB 3;  Length 2042;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          ||||||  ||||  | ||||||     |||||     |  |  |  |||||  |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || ||||  |  |   |  | | | || |  |  |  |||||  |     |||  || |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || ||| ||||| || ||  ||  |  |  | | | | ||| |||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  ||||| || ||  |  || ||| | || || ||| ||  |  |  |  |  | ||||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  |  | |  |  |||  |||         |  |  |  |  |  |  ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || ||||  |||||  |||  |||||||| || ||    |||||||  |  |  |||  |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |  ||  |  | | |  ||  || || |  |  |  |  |  |  |  |  |  |  |||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
          |  |   |  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || |||||||| || |||||  |     |  |  |  |  |  |||  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || |||||  ||   ||  |||||  |||  |  |  |  |  |  |  |  |
```

```
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    ||||  ||  | ||     || ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |      ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 9
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             5.9%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || ||||  |   |   | | |    | | || |      | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |   ||  |  | |  |     |    ||          | |  |  |  |  ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
          | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || ||||||||  || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    |||||  |||  |   | |||  | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |      ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 10
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
```

```
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match                 5.9%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity  53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          ||||| ||||  | |||||||      |||||   |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          |  || ||||  |   |   |  | ||  |     | |||||  |            |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || ||| ||||| || ||  ||  ||  |  |  |  | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  |||||  || ||  | || ||| | || || || |||  |||  || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  |  |  |  |  ||         |  |  |  |  |  ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          ||  |||||  |||||  |||  ||||||||  || ||  |||||||      || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |  ||     |  |  |  |  ||  ||  |  |  |  |  |  |  |  |  |  |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
          |  |  |  ||||  ||  ||  |  ||  || ||  ||  |  ||  ||  |||||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||  || |||||||| || |||||  |    ||  ||  |  |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||  ||  |||||  ||||  |  |  ||| |  |  ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          |||||  || ||||| ||  ||  |  |||  ||  ||  |||  |  ||  |||| |||
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  ||  |  |||||  || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044

RESULT 11
US-08-273-411-4
, Sequence 4, Application US/08273411
, Patent No. 5625124
,  GENERAL INFORMATION:
;     APPLICANT:  Falk, Per
;     APPLICANT:  Gordon, Jeffrey I.
;     TITLE OF INVENTION:  Animal Model for Gastro-Intestinal
;     TITLE OF INVENTION:  Disease
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Patrea L. Pabst
;       STREET:  1100 Peachtree Street, Suite 2800
;       CITY:  Atlanta
;       STATE:  Georgia
;       COUNTRY:  USA
;       ZIP:  30309-4530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/273,411
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pabst, Patrea L.
;       REGISTRATION NUMBER:  31,284
;       REFERENCE/DOCKET NUMBER:  WU106
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (404) 815-6508
;       TELEFAX:  (404) 815-6555
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2043 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..361
;       OTHER INFORMATION:  /note= "Nucleotides 73 through 1158 encode the GDP-L-fuco
;     PUBLICATION INFORMATION:
;       AUTHORS:  Kukowska-Latallo, et al.
;       JOURNAL:  Genes & Development
;       VOLUME:  4
;       PAGES:  1288-1303
;       DATE:  1990
;       RELEVANT RESIDUES IN SEQ ID NO:  4: FROM 1 TO 2043
US-08-273-411-4

  Query Match                 5.9%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity  53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | |||||||      |||||   |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          |  || ||||  |   |   |  | ||  |     | |||||  |            |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || ||| ||||| || ||  ||  ||  |  |  |  | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  |||||  || ||  | || ||| | || || || |||  |||  || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |  ||  |  |  |  |  ||         |  |  |  |  |  ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642
```

```
Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |    ||          | |  | |  ||   || || |    |    |     |  | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
            | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||   |    ||||  ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 12
US-08-525-058A-1
; Sequence 1, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,058A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO

US-08-525-058A-1

Query Match 5.9%; Score 128.2; DB 1; Length 2043;
Best Local Similarity 53.0%; Pred. No. 9.5e-17;
Matches 374; Conservative 0; Mismatches 313; Indels 18; Gaps 4;

```
Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
          | || ||||  |   |   |  | | |  | | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
          |||||  ||| | || || |||  ||||| || ||  ||  ||  |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
          |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
          |   ||  |  | | |  |    ||          | |  |  |  |  ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
          |    ||          | |  | |  ||   || || |    |    |     |  | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
            | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
          ||||| || ||||| || ||   |    ||||  ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 13
US-08-696-731-1
; Sequence 1, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.

```
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/696,731
;       FILING DATE:  14-AUG-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2043 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     ANTI-SENSE:  NO
US-08-696-731-1

  Query Match                 5.9%;  Score 128.2;  DB 2;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
           ||||| |||| | ||||||      |||||    |  |  | ||||| |  | ||||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501
           | || ||||  |   |   |  | | |  | | || |     | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561
           |||||| ||| | || || ||| ||||| || ||  ||  ||    |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621
           |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672
           |   ||  | | |   |   ||           | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732
           || |||||  |||||  |||  ||||||||| || ||    ||||||||     || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792
           |   ||          | |  |  |  ||  || | |   |   |     |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852
           |  |    |  ||||    ||  | ||     |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909
           ||  | |||   || ||||||||| || ||||| |     || ||  |  |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969
           || || || || ||||| ||     ||     |||||  ||||  |   | |||  | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029
           |||||  ||  |||||  ||  ||   |     ||||  ||  |  ||     || |||  ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
           |||||  ||  |        |||||  ||  ||||||  ||||  |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-09-042-531-1
; Sequence 1, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-09-042-531-1

  Query Match                 5.9%;  Score 128.2;  DB 4;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444
```

```
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 15
PCT-US91-00899-1
; Sequence 1, Application PC/TUS9100899
;  GENERAL INFORMATION:
;    APPLICANT:  Lowe, John B.
;    TITLE OF INVENTION:  Method and Products For the Synthesis of
;    TITLE OF INVENTION:  Oligosaccharide Structures on Glycoproteins, Glycolipids,
;    TITLE OF INVENTION:  or as Free Molecules, and For the Isolation of Cloned
;    TITLE OF INVENTION:  Genetic Sequences That Determine These Structur
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00899
;      FILING DATE:  19910214
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye Ph.D., Jean-Paul
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-021-55 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-5940
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Blood
;      CELL LINE:  A431
PCT-US91-00899-1

  Query Match             5.9%;  Score 128.2;  DB 5;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 9.5e-17;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    385 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 444

Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    445 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 501

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    502 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 561

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    562 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 621

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    622 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 672

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    673 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 732

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    733 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 792

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    793 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 852

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    853 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 909

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    910 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 969

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    970 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 1029

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1030 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074
```

```
                ||||||  ||  |        ||||| ||  ||||||  ||||  |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:20:16
Job time : 54.2762 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 1353.36 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-3_COPY_115_1194
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1080	100.0	2139	10	AB015426	AB015426 Mus muscu
2	1008	93.3	1128	10	AF345993	AF345993 Rattus no
3	1004.8	93.0	2156	10	AB049819	AB049819 Rattus no
4	974.4	90.2	1705	10	AF230460	AF230460 Cricetulu
5	936	86.7	2501	9	HSA238701	AJ238701 Homo sapi
6	932.8	86.4	80247	9	AL512406	AL512406 Human DNA
7	931.2	86.2	3019	9	AB023021	AB023021 Homo sapi
8	740	68.5	71396	2	AC100674	AC100674 Mus muscu
9	727.6	67.4	1707	5	AB035906	AB035906 Gallus ga
10	675.2	62.5	1080	5	AB035905	AB035905 Xenopus l
11	268.2	24.8	1567	5	AB023627	AB023627 Danio rer
12	228.2	21.1	1832	5	AB023628	AB023628 Danio rer
13	220.6	20.4	1264	5	AF288369	AF288369 Gallus ga
14	206.6	19.1	1092	5	AB035907	AB035907 Xenopus l
15	175.2	16.2	2657	5	GGU73678	U73678 Gallus gall
16	157	14.5	1429	10	CGU78737	U78737 Cricetulus
17	157	14.5	1462	10	AF090450	AF090450 Cricetulu
18	155.4	14.4	1718	10	AF090449	AF090449 Cricetulu
c 19	145.6	13.5	71396	2	AC100674	AC100674 Mus muscu
20	135.6	12.6	1044	5	AB035908	AB035908 Gallus ga
21	134.8	12.5	2293	4	BTA134FTG	X87810 B.taurus al
22	134.8	12.5	12607	4	BTA132772	AJ132772 Bos tauru
23	134.2	12.4	1158	10	AB039126	AB039126 Mus muscu
24	133.4	12.4	1113	9	AF345884	AF345884 Saimiri s
25	133	12.3	1119	9	AF345881	AF345881 Macaca mu
26	132.6	12.3	1170	10	AB039127	AB039127 Mus muscu
27	132.6	12.3	1170	10	AB039131	AB039131 Mus muscu
28	132.6	12.3	1615	10	MMU33457	U33457 Mus musculu
29	131.4	12.2	1170	10	AB039133	AB039133 Mus spici
30	131	12.1	1170	10	AB039124	AB039124 Mus muscu
31	131	12.1	1170	10	AB039125	AB039125 Mus muscu
32	131	12.1	1170	10	AB039128	AB039128 Mus muscu
33	131	12.1	1170	10	AB039129	AB039129 Mus muscu
34	131	12.1	1170	10	AB039130	AB039130 Mus muscu
35	131	12.1	1170	10	AB039132	AB039132 Mus muscu
36	130.8	12.1	1302	10	AB049938	AB049938 Rattus no
37	130.8	12.1	169446	2	AC097778	AC097778 Rattus no
38	130	12.0	1613	4	BTA132776	AJ132776 Bos tauru
39	130	12.0	1619	4	BTA132775	AJ132775 Bos tauru
40	130	12.0	1644	4	BTA132774	AJ132774 Bos tauru
41	130	12.0	1873	4	BTA132773	AJ132773 Bos tauru
42	129.8	12.0	1162	9	PTY14033	Y14033 Pan troglod
c 43	129.8	12.0	80893	9	AC024592	AC024592 Homo sapi
44	129.8	12.0	187488	2	AC010454	AC010454 Homo sapi
c 45	128.8	11.9	169893	2	AC021165	AC021165 Homo sapi

ALIGNMENTS

RESULT 1
AB015426
LOCUS AB015426 2139 bp mRNA linear ROD 23-JUN-1999
DEFINITION Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete cds.
ACCESSION AB015426
VERSION AB015426.1 GI:3702718
KEYWORDS Fut9; alpha1,3-fucosyltransferase IX.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K. and Narimatsu,H.
TITLE Expression cloning and characterization of a novel murine alpha1, 3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15) epitope in brain and kidney
JOURNAL J. Biol. Chem. 273 (41), 26729-26738 (1998)
MEDLINE 98434588
REFERENCE 2 (bases 1 to 2139)
AUTHORS Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN

  Query Match             100.0%;  Score 1080;  DB 10;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-278;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 174

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 234

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 294

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 354

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 414

Qy     301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 474

Qy     361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 534

Qy     421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 594

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 654

Qy     541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 714

Qy     601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 774

Qy     661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 834

Qy     721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 894

Qy     781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 954

Qy     841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 1014

Qy     901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074

Qy     961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1134

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194


RESULT 2
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
```

```
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN

  Query Match           93.3%;  Score 1008;  DB 10;  Length 1128;
  Best Local Similarity  95.8%;  Pred. No. 5.7e-259;
  Matches 1035;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     49 ATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGGGC 108

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          |||||| |||||||||| || || || ||||||||||||||||||||||||||||||||||
Db    109 TGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 168

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          |||||||||||||| |||||||||||||||||||||||||||||  ||||||||||||
Db    169 ATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATTAT 228

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||||||||| ||||||||||||||||||||||||||||||
Db    229 TTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 288

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||| |||||||| |||||||||||||||||||||||| |||||||| ||||||
Db    289 TCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGTAC 348

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db    349 AACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTA 408

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db    409 CCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCACT 468

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    469 CACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCCGT 528

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          |||||||||| |||||||||| |||||||||||||||||||||  ||||||||||||
Db    529 GATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTT 588

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| |||||||| |||||||||||||||||||||||| || ||||||
Db    589 GAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCAT 648

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    649 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGCAA 708

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||| |||||||| ||||| |||||||||||||||||||||||||||||||||||||||
Db    709 GCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 768

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    769 TATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 828

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    829 TTTTTGGCTGGTTCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 888

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    889 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTGGCAAAA 948

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||||||||| |||||||||||||||||||| || |||||||||| |||||||| |||

Db    949 TATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGAAG 1008

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db   1009 GATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATCAT 1068

Qy   1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1069 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1128


RESULT 3
AB049819
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN

  Query Match           93.0%;  Score 1004.8;  DB 10;  Length 2156;
  Best Local Similarity  95.6%;  Pred. No. 4.1e-258;
  Matches 1033;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    346 ATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGGGC 405

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          |||||| |||||||||| || || || ||||||||||||||||||||||||||||||||||
Db    406 TGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 465
```

```
Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          |||||||||||||| |||||||||||||||||||||||||||||||  ||||||||||||
Db    466 ATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATTAT 525

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          |||||  ||||||||  ||||||||||||||  ||||||||||||||||||||||||||||
Db    526 TTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 585

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          ||||||||||| ||||||||  |||||||||||||||||||||||| ||||||||  ||||||
Db    586 TCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGTAC 645

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||||||||||||| |||||  |||||||||||||||||||||||||||||
Db    646 AACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTA 705

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||||
Db    706 CCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCACT 765

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    766 CACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCCGT 825

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          |||||||||| |||||||||| |||||||||||||||||||||  ||||||||||||||
Db    826 GATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTT 885

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| |||||||| |||||||||||||||||||||||| || ||||||||
Db    886 GAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCAT 945

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    946 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGCAA 1005

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          |||||  ||||||||  |||||  |||||||||||||||||||||||||||||||||||||
Db   1006 GCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 1065

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1066 TATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 1125

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          |||||||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||
Db   1126 TTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1185

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||||||
Db   1186 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGTGAGTTGGCAAAA 1245

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||||||| |||||||||||||||||||| || |||||||||| ||||||||  |||
Db   1246 TATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGAAG 1305

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||  |||
Db   1306 GATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATCAT 1365

Qy   1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   1366 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1425
```

```
RESULT 4
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN

  Query Match             90.2%;  Score 974.4;  DB 10;  Length 1705;
  Best Local Similarity   93.9%;  Pred. No. 5.6e-250;
  Matches 1014;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          |||||||||||||||||||||||||||||||||| ||||| ||||||||||  | ||||||| |
Db     62 ATGACATCAACATCCAAAGGCATTCTTCGTCCATTCCTAATCGTCTGTGTTATCCTGGCC 121

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          |||||||||||||||||||  ||||| |||||||||||||||||||||| |||||||||
Db    122 TGCTTCATGGCATGTCTGCTTATTTACATCAAGCCCACCAACAGCTGGGTGTTCAGTCCA 181

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 ATGGAGTCTGCCAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 241

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| ||||| || ||||| ||||||||  |||||||||||||||||||||||||||||||
Db    242 TTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 301

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||  |||||||| |||||||||||||||||||  |||||||||| |||||
Db    302 TCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACAGACCGCTCACTGTAC 361

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||||||||||||| |||||  |||||||||||||||||||||||||||||
Db    362 AACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTA 421

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||||||  ||||| |||||||||||||||||||||||||||  ||||||| ||||||
```

```
Db    422 CCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAACCTAGAGTCTCCCACT 481

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          || ||||||||||||||||||||| || |||||||||||||||||| |||||||||
Db    482 CATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACTCTAACATATCGCCGT 541

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 601

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| |||||| |||||||||||||||||||||||| || ||||||
Db    602 GAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAACTGGAATCCGGAGCAT 661

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          |||||||||||||||| || |||||||||||||||||||||||| |||||||| |||
Db    662 GCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATCCATACCTATGGGCAA 721

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          || || |||||||||||||| || |||||||| ||||||||||||||||||||||||||
Db    722 GCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATATCTACTTGTAAATTT 781

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          |||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    782 TATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAAAAGCTCTACAATGCA 841

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          |||||||||| |||||||||||||| ||||| ||||| ||||||||||||||||||||||
Db    842 TTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAAAACTATGAGAATTAT 901

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||| |||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    902 ATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGTTGGCAAAA 961

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          ||||| |||||||||||||||||||||||||||||||||||||| || |||||||| |||
Db    962 TATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTATTTCAACTGGAGGAAG 1021

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          ||||| |||||||||||| ||||| |||||||||||||||||||||||||| || || ||
Db   1022 GATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGCCTGGCTTGTGATCAC 1081

Qy   1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          |||||||||||| ||||| |||||||||| ||||||||||||||||||||||||||||||
Db   1082 GTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAATGGTTTTGGAATTAA 1141
```

```
RESULT 5
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN
```

```
  Query Match             86.7%;  Score 936;  DB 9;  Length 2501;
  Best Local Similarity   91.7%;  Pred. No. 1.1e-239;
  Matches  990;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          |||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db    105 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 164

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    165 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 224

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db    225 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 284

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db    285 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 344

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db    345 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 404

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||| || ||||| || ||  ||||||||||||||||||||||||||||||
Db    405 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGCTGGGATCTGACTAACTTA 464

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||| || |||||||||| ||||||||||||||||||||||||| || ||||| |||
Db    465 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 524

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||| |||||||||||||||||||| |||||||| |||||||||| ||||| ||||||
Db    525 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 584

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     585 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 644

Qy     541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
           |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db     645 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 704

Qy     601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
           ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db     705 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 764

Qy     661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
           ||||| |||||||| || || ||||||||| ||||||| |||||||| ||||||||||||
Db     765 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 824

Qy     721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
           |||||||| |||||||| ||||| ||||| |||||||||| |||||||| ||||||||
Db     825 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 884

Qy     781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
           ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||||||||||
Db     885 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 944

Qy     841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
           |||||||| |||||||||||||||||||||||| |||||||||||||| |||||||
Db     945 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1004

Qy     901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||||||||||| |||||||||||||||| ||||||||||||||||||||||||| |||
Db    1005 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1064

Qy     961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
           ||||| |||||||| || ||||| ||||||||||||||||||  |||| ||||| |||
Db    1065 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1124

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1125 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1184
```

```
RESULT 6
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-504J9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-381A2 is at 80148 in this sequence.
            The true right end of clone RP11-77B15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .80247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-504J9"
                     /clone_lib="RPCI-11.2"
     repeat_region   184. .1422
                     /note="L2 repeat: matches 875. .2419 of consensus"
     repeat_region   1744. .1834
                     /note="MIR repeat: matches 166. .258 of consensus"
     repeat_region   2950. .2977
                     /note="14 copies 2 mer ca 100% conserved"
     repeat_region   3411. .3669
                     /note="LTR16C repeat: matches 93. .364 of consensus"
     repeat_region   4372. .4704
                     /note="MLT1A1 repeat: matches 5. .365 of consensus"
     misc_feature    complement(7821. .8406)
                     /note="match: GSS: Em:AQ550748"
     repeat_region   8126. .8184
                     /note="MER58B repeat: matches 272. .332 of consensus"
     repeat_region   8185. .8696
                     /note="MLT2D repeat: matches 1. .513 of consensus"
     repeat_region   8779. .8817
                     /note="MLT2D repeat: matches 504. .542 of consensus"
     repeat_region   8819. .8974
                     /note="MER58A repeat: matches 6. .162 of consensus"
     misc_feature    complement(9443. .9891)
                     /note="match: GSS: Em:B90633"
     misc_feature    complement(9452. .9883)
                     /note="match: GSS: Em:AQ119695"
     repeat_region   9782. .9894
                     /note="MIR repeat: matches 92. .205 of consensus"
     repeat_region   10089. .10212
                     /note="MIR repeat: matches 135. .262 of consensus"
     misc_feature    complement(10093. .10671)
                     /note="match: GSS: Em:AQ237061"
     repeat_region   10223. .10282
                     /note="MIR repeat: matches 203. .262 of consensus"
     misc_feature    complement(10315. .10631)
                     /note="match: GSS: Em:AQ200259"
     misc_feature    10805. .11186
                     /note="match: GSS: Em:AQ072353"
     repeat_region   11657. .11844
                     /note="L2 repeat: matches 2546. .2750 of consensus"
     misc_feature    complement(11659. .12084)
                     /note="match: GSS: Em:AQ518361"
     misc_feature    complement(11674. .11987)
                     /note="match: GSS: Em:AQ094678"
     misc_feature    12083. .12588
                     /note="match: GSS: Em:AQ691235"
     repeat_region   12596. .12810
                     /note="MIR repeat: matches 23. .248 of consensus"
     repeat_region   12920. .12986
                     /note="L1MA5 repeat: matches 6228. .6295 of consensus"
     repeat_region   12987. .13169
                     /note="L1MA5 repeat: matches 5852. .6043 of consensus"
```

```
repeat_region   13168. .13280
                /note="L1MA8 repeat: matches 5710. .5832 of consensus"
repeat_region   13851. .17218
                /note="L1PB1 repeat: matches 2787. .6155 of consensus"
repeat_region   18903. .18953
                /note="MLT2CB repeat: matches 439. .501 of consensus"
repeat_region   18984. .19433
                /note="MLT2CB repeat: matches 1. .460 of consensus"
misc_feature    complement(19108. .19690)
                /note="match: GSS: Em:AQ563935"
repeat_region   20683. .20801
                /note="AluJo/FRAM repeat: matches 169. .288 of consensus"
repeat_region   20802. .20862
                /note="AluJ/FLAM repeat: matches 3. .66 of consensus"
repeat_region   20912. .21033
                /note="MIR repeat: matches 114. .237 of consensus"
repeat_region   21047. .21128
                /note="41 copies 2 mer ta 63% conserved"
repeat_region   21676. .21859
                /note="L2 repeat: matches 2130. .2308 of consensus"
repeat_region   21890. .23612
                /note="L1P repeat: matches 841. .2570 of consensus"
repeat_region   23601. .26609
                /note="L1PA5 repeat: matches 3107. .6145 of consensus"
repeat_region   26620. .26789
                /note="MER5A repeat: matches 1. .174 of consensus"
repeat_region   28521. .28627
                /note="MER39 repeat: matches 5. .111 of consensus"
repeat_region   28656. .28837
                /note="MER39b repeat: matches 54. .235 of consensus"
repeat_region   28838. .29142
                /note="AluSc repeat: matches 4. .308 of consensus"
repeat_region   29143. .29494
                /note="MER39b repeat: matches 235. .549 of consensus"
repeat_region   30529. .31344
                /note="L1PA3 repeat: matches 5331. .6146 of consensus"
repeat_region   32234. .32340
                /note="L2 repeat: matches 2627. .2750 of consensus"
repeat_region   32399. .32578
                /note="MIR repeat: matches 12. .203 of consensus"
repeat_region   32816. .32979
                /note="MIR repeat: matches 91. .259 of consensus"
misc_feature    complement(33006. .33450)
                /note="match: GSS: Em:AQ261227"
misc_feature    complement(33058. .33444)
                /note="match: GSS: Em:AQ040453"
repeat_region   33312. .33596
                /note="L1M4 repeat: matches 3473. .3773 of consensus"
repeat_region   34856. .34915
                /note="30 copies 2 mer aa 68% conserved"
misc_feature    complement(34987. .35547)
                /note="match: GSS: Em:AQ321598
                match: STS: Em:G55571"
misc_feature    complement(36488. .36936)
                /note="match: GSS: Em:B41073"
misc_feature    complement(36654. .37090)
                /note="match: GSS: Em:AQ772584"
misc_feature    37102. .37591
                /note="match: GSS: Em:AQ489567"
misc_feature    complement(37581. .38221)
                /note="match: GSS: Em:AQ418233"
repeat_region   37720. .37772
                /note="L2 repeat: matches 2657. .2710 of consensus"
misc_feature    complement(37778. .38215)
                /note="match: GSS: Em:AQ819773"
misc_feature    38218. .38732
                /note="match: GSS: Em:AQ599868"
repeat_region   38861. .39164
                /note="AluSx repeat: matches 1. .305 of consensus"
repeat_region   39612. .39653
                /note="21 copies 2 mer ac 90% conserved"
repeat_region   40854. .40909
                /note="L2 repeat: matches 2655. .2710 of consensus"
repeat_region   42169. .42270
                /note="MER5A repeat: matches 1. .111 of consensus"
repeat_region   42411. .42548
                /note="MER5B repeat: matches 21. .181 of consensus"
repeat_region   42780. .42938
                /note="MER5B repeat: matches 17. .177 of consensus"
repeat_region   43229. .43326
                /note="MIR repeat: matches 40. .148 of consensus"
repeat_region   45383. .45428
                /note="MER5B repeat: matches 18. .62 of consensus"
misc_feature    complement(47284. .47971)
                /note="match: STS: Em:G51995
                match: GSS: Em:AQ269059"
misc_feature    complement(47537. .48011)
                /note="match: STS: Em:G36899"
misc_feature    complement(47723. .48014)
                /note="match: STS: Em:G30049"
misc_feature    47972. .48451
                /note="match: GSS: Em:AQ336527"
misc_feature    complement(48518. .48985)
                /note="match: GSS: Em:AQ801834"
repeat_region   48597. .48803
```

Query Match 86.4%; Score 932.8; DB 9; Length 80247;
Best Local Similarity 91.5%; Pred. No. 8.5e-239;
Matches 988; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db  35559 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 35618

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db  35619 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 35678

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db  35679 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 35738

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db  35739 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 35798

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db  35799 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 35858

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          ||||||||||||| || ||||| || ||  |||||||| |||||||||| || |||
Db  35859 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 35918

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||| || |||||||||| |||||||||||||||||||| || ||||| |||
Db  35919 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 35978

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||| |||||||||||||||||| |||||||| |||||||||| ||||| ||||||
Db  35979 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 36038

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db  36039 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 36098

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db  36099 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 36158

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db  36159 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 36218

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||| |||||||| || || ||||||||| ||||||| ||||||||||||||||||||
```

```
Db  36219 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTACTTGTAAATTT 36278

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          |||||||| |||||||| ||||| ||||| |||||||||||| |||||||| ||||||||
Db  36279 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 36338

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          ||| ||||||| || |||||||||||| ||||| ||||||||||||||||||||||||||
Db  36339 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 36398

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||| |||||||||||||||||||||||||| ||||||||||||||| |||||||
Db  36399 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 36458

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||||||||||| |||||||||||||||| |||||||||||||||||||||||| |||
Db  36459 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 36518

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          ||||| |||||||| || ||||| ||||||||||||||||||  |||| ||||| |||
Db  36519 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 36578

Qy   1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36579 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 36638


RESULT 7
AB023021
LOCUS       AB023021                3019 bp    mRNA    linear   PRI 23-JUN-1999
DEFINITION  Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB023021
VERSION     AB023021.1  GI:5139692
KEYWORDS    alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S.,
            Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and
            Narimatsu,H.
  TITLE     Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved
            between human and mouse; molecular cloning, characterization and
            tissue distribution of human Fuc-TIX
  JOURNAL   FEBS Lett. 453, 237-242 (1999)
REFERENCE   2  (bases 1 to 3019)
  AUTHORS   Kaneko,M., Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
ORIGIN

  Query Match           86.2%;  Score 931.2;  DB 9;  Length 3019;
  Best Local Similarity  91.4%;  Pred. No. 2.1e-238;
  Matches  987;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          |||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db    295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db    415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db    475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db    535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||| || ||||| || ||  |||||||||| |||||||||| || |||
Db    595 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 654

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||| || |||||||||| |||||||||||||||||||||||| || ||||| |||
Db    655 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 714

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||| |||||||||||||||||| |||||||| |||||||||| |||| ||||||
Db    715 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 774

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db    775 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db    835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db    895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||| |||||||| || || |||||||| ||||||| |||||||| |||||||||||||
Db    955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          |||||||| |||||||| ||||| ||||| |||||||||||| |||||||| ||||||||
Db   1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 1074

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          ||| ||||||| || |||||||||||| ||||| ||||||||||||||||||||||||||
Db   1075 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1134

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||| |||||||||||||||||||||||||| |||||||||||||||| |||||||
Db   1135 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1194

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||||||||||| |||||||||||||||| |||||||||||||||||||||||| |||
Db   1195 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
```

```
          ||||| |||||||| || ||||| ||||||||||||||||||||  |||| ||||| |||
Db    1255 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1314

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1374


RESULT 8
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1702: contig of 814 bp in length
            *     1703 1802: gap of      100 bp
            *     1803     2642: contig of 840 bp in length
            *     2643 2742: gap of      100 bp
            *     2743     3637: contig of 895 bp in length
            *     3638 3737: gap of      100 bp
            *     3738     4635: contig of 898 bp in length
            *     4636 4735: gap of      100 bp
            *     4736     5657: contig of 922 bp in length
            *     5658 5757: gap of      100 bp
            *     5758     6692: contig of 935 bp in length
            *     6693 6792: gap of      100 bp
            *     6793     7703: contig of 911 bp in length
            *     7704 7803: gap of      100 bp
            *     7804     8742: contig of 939 bp in length
            *     8743 8842: gap of      100 bp
            *     8843     9787: contig of 945 bp in length
            *     9788 9887: gap of      100 bp
            *     9888    10819: contig of 932 bp in length
            *    10820 10919: gap of      100 bp
            *    10920    11808: contig of 889 bp in length
            *    11809 11908: gap of      100 bp
            *    11909    12830: contig of 922 bp in length
            *    12831 12930: gap of      100 bp
            *    12931    13847: contig of 917 bp in length
            *    13848 13947: gap of      100 bp
            *    13948    14886: contig of 939 bp in length
            *    14887 14986: gap of      100 bp
            *    14987    15911: contig of 925 bp in length
            *    15912 16011: gap of      100 bp
            *    16012    16903: contig of 892 bp in length
            *    16904 17003: gap of      100 bp
            *    17004    17906: contig of 903 bp in length
            *    17907 18006: gap of      100 bp
            *    18007    18931: contig of 925 bp in length
            *    18932 19031: gap of      100 bp
            *    19032    19945: contig of 914 bp in length
            *    19946 20045: gap of      100 bp
            *    20046    20951: contig of 906 bp in length
            *    20952 21051: gap of      100 bp
            *    21052    21949: contig of 898 bp in length
            *    21950 22049: gap of      100 bp
            *    22050    22950: contig of 901 bp in length
            *    22951 23050: gap of      100 bp
            *    23051    23945: contig of 895 bp in length
            *    23946 24045: gap of      100 bp
            *    24046    24955: contig of 910 bp in length
            *    24956 25055: gap of      100 bp
            *    25056    25980: contig of 925 bp in length
            *    25981 26080: gap of      100 bp
            *    26081    27052: contig of 972 bp in length
            *    27053 27152: gap of      100 bp
            *    27153    28067: contig of 915 bp in length
            *    28068 28167: gap of      100 bp
            *    28168    29076: contig of 909 bp in length
            *    29077 29176: gap of      100 bp
            *    29177    30093: contig of 917 bp in length
            *    30094 30193: gap of      100 bp
            *    30194    31099: contig of 906 bp in length
            *    31100 31199: gap of      100 bp
            *    31200    32097: contig of 898 bp in length
            *    32098 32197: gap of      100 bp
            *    32198    33113: contig of 916 bp in length
            *    33114 33213: gap of      100 bp
            *    33214    34125: contig of 912 bp in length
            *    34126 34225: gap of      100 bp
            *    34226    35133: contig of 908 bp in length
            *    35134 35233: gap of      100 bp
            *    35234    36150: contig of 917 bp in length
            *    36151 36250: gap of      100 bp
            *    36251    37171: contig of 921 bp in length
            *    37172 37271: gap of      100 bp
```

```
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp

  Query Match             68.5%;  Score 740;  DB 2;  Length 71396;
  Best Local Similarity   83.4%;  Pred. No. 3.4e-187;
  Matches  844;  Conservative    0;  Mismatches  154;  Indels   14;  Gaps    6;

Qy      72 ATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGC 131
           || |||    | | |      |  |||||||||||||| ||||||||||||||||||||||
Db   65456 ATCTCTCGCCCTGTGGTGGATTCCCACCAACAGCTCGGTCTTCAGTCCAATGGAGTCTGC 65515

Qy     132 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65516 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 65575

Qy     192 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65576 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 65635

Qy     252 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65636 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 65695

Qy     312 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65696 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 65755

Qy     372 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65756 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 65815

Qy     432 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65816 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 65875

Qy     492 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 551
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65876 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 65935

Qy     552 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 611
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65936 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 65995

Qy     612 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 671
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65996 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 66055

Qy     672 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATT 731
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   66056 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTC-TC 66114

Qy     732 TGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGG 791
           |||||||||| |    | |     |  | ||        | ||     ||||||||||
Db   66115 TGAAAACTCATTCACAAGATTACTCCCAAAAG--------TCCACATGCTTTTTGGCTGG 66166

Qy     792 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66167 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 66226

Qy     852 TTCATTCATTC--ATGTGGAAGATTTTAACTCT-CCCAGTGAGTTAGCAAAATATCTGAA 908
           |||||||||||  ||||||||||| ||||| |||||||||||||||    |||||
Db   66227 TTCATTCATTCCATCGTGGAAGATTTTCACTCTCCCCAGTGAGTTAGCAAATATCCTGAA 66286

Qy     909 GGAAGTTGACAAAAACAATAAGTT-GTACCTTAGTTACTTTAACT-GGAGAAAGGATTTT 966
           |||||| ||||| |||||||||| |||||||||||| ||||| | |||||||||| |
Db   66287 GGAAGTCGACAANAACAATAAGTTGGTACCTTAGTTACTCTAACTCGNAGAAAGGATTCT 66346

Qy     967 ACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAA 1026
           || |
Db   66347 ACCGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 66406

Qy    1027 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1078
                                                           |||
Db   66407 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGATT 66458

RESULT 9
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
```

```
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN
```

Query Match 67.4%; Score 727.6; DB 5; Length 1707;
Best Local Similarity 79.5%; Pred. No. 6.5e-184;
Matches 859; Conservative 1; Mismatches 220; Indels 0; Gaps 0;

```
Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
Db    268 ATGACATCAACATCTAAAGGAATTTTCCGGCCATTTTTTATTGTCTTCATTGTYCTTGGT 327

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
Db    328 TGTTTCATGGCATTTATACTAATTTATATTAAACCAACAAGCAGCTGGATCTCTAATCCT 387

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
Db    388 GTGGAATCAGCCAGCTCAGTTTTGAAAATGAAGAGCTTCTTTTCTTCCAAAACTGATAAT 447

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
Db    448 TTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGACATTCGATCTAACA 507

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
Db    508 TCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTGACCGCTCACTATAT 567

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
Db    568 AACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGGATCTGACTAATTTA 627

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
Db    628 CCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACTTGGAATCTCCAACT 687

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
Db    688 CATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCCTGACTTACCGGCGT 747

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
Db    748 GATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAGGCGCCTTTACATAT 807

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
Db    808 GAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACTGGAACCCTGAGCAT 867

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
Db    868 GCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCCATACCTACGGGCAA 927

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
Db    928 GCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCTCCAGTTGCAAATTC 987

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
Db    988 TACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGAAACTTTACAATGCT 1047

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
Db   1048 CTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAAACTACGAGAATTAC 1107

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
Db   1108 ATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAAGAGAGCTTGCAGAA 1167

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
Db   1168 TATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACTTCAACTGGAGGAAG 1227

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
Db   1228 GATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTCTTGCTTGTGATCAT 1287

Qy   1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
Db   1288 GTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAATGGTTTTGGAACTAA 1347
```

```
RESULT 10
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
```

```
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN

  Query Match            62.5%;  Score 675.2;  DB 5;  Length 1080;
  Best Local Similarity  76.6%;  Pred. No. 6.9e-170;
  Matches  827;  Conservative    0;  Mismatches  253;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
Db       1 ATGACATCAGCATCTAAAGGAATCTTACGGGCATTTTTAGCTGTCAGCGGTGTGTTGATC 60

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
Db      61 TGTTTTATGATATGTTTACTGATATACGTCAAACCAACAAACAACTGGATTTCCAGTCCT 120

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
Db     121 ATCGAATCAGCACATTCAGTCCTTAAAATGAAGAATTTTTTCTCCTCTAAACAGAATTAT 180

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
Db     181 TACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGACTTTTGAACTGAAA 240

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
Db     241 TCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTGACCGCGCACTCTAC 300

Qy     301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
Db     301 AACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGGATCTCATTAACCTG 360

Qy     361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
Db     361 CCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATCTGGAATCACCAACA 420

Qy     421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
Db     421 CATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTCTGACTTACAGGCGT 480

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
Db     481 GACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAAAACCATTTGACTTT 540

Qy     541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
Db     541 GAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACTGGAACCCAGAACAT 600

Qy     601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
Db     601 GCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTATAACTTATGGGCAA 660

Qy     661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
Db     661 GCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTGCATCTTGCAAATTT 720

Qy     721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
Db     721 TATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAAAACTTTATAATGCA 780

Qy     781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
Db     781 TTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAAACTATGAGAACTAC 840

Qy     841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
Db     841 ATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAAGAGAATTAGCTGAC 900

Qy     901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
Db     901 CATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACTACAACTGGAGGAAA 960

Qy     961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
Db     961 CATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCTTAGCGTGTGACCAT 1020

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
Db    1021 GTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGTGGTTCTGGAACTAA 1080


RESULT 11
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN

  Query Match            24.8%;  Score 268.2;  DB 5;  Length 1567;
  Best Local Similarity  57.8%;  Pred. No. 6.1e-61;
  Matches  521;  Conservative    0;  Mismatches  368;  Indels   12;  Gaps    2;

Qy     183 TAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATC 242
```

```
Db    461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy    243 CTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAA 302

Db    521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy    303 CAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACC 362

Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGGAT---------TTGCC 631

Qy    363 TCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCA 422

Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy    423 CACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGA 482

Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy    483 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTT 539

Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy    540 TGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 599

Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy    600 TGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCA 659

Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy    660 AGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATT 719

Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy    720 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 779

Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy    780 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 839

Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy    840 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAA 899

Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy    900 ATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAA 959

Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy    960 GGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCA 1019

Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy   1020 TGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1079

Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy   1080 A 1080

Db   1352 A 1352


RESULT 12
AB023628
LOCUS       AB023628               1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match             21.1%;  Score 228.2;  DB 5;  Length 1832;
  Best Local Similarity   55.0%;  Pred. No. 3.1e-50;
  Matches  496;  Conservative    0;  Mismatches  393;  Indels   12;  Gaps    2;

Qy    183 TAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATC 242

Db    619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy    243 CTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAA 302

Db    679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy    303 CAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACC 362

Db    739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGCACAGAT---------TTTCC 789

Qy    363 TCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCA 422

Db    790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy    423 CACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGA 482

Db    850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy    483 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA---AATCCCTTTGTGTT 539

Db    910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy    540 TGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 599

Db    970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy    600 TGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCA 659
```

```
Db   1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy    660 AGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATT 719

Db   1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy    720 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 779

Db   1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy    780 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 839

Db   1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy    840 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAA 899

Db   1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy    900 ATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAA 959

Db   1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy    960 GGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCA 1019

Db   1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy   1020 TGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1079

Db   1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy   1080 A 1080

Db   1510 A 1510
```

```
RESULT 13
AF288369
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
```

```
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN

  Query Match           20.4%;  Score 220.6;  DB 5;  Length 1264;
  Best Local Similarity 57.2%;  Pred. No. 3.4e-48;
  Matches  441;  Conservative    0;  Mismatches  324;  Indels    6;  Gaps    2;

Qy    193 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 252

Db    175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy    253 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 312

Db    235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy    313 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 372

Db    295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy    373 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 432

Db    355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy    433 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 492

Db    415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy    493 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 552

Db    475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy    553 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 612

Db    535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy    613 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 672

Db    595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGCGGCAACAC---TTG 651

Qy    673 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 732

Db    652 CCCCTGTCCATGGACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy    733 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTAC---AATGCATTTTTGGCT 789

Db    712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy    790 GGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCT 849

Db    772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy    850 GATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAG 909

Db    832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy    910 GAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db    892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942
```

```
RESULT 14
AB035907
LOCUS       AB035907               1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
```

```
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN

  Query Match           19.1%;  Score 206.6;  DB 5;  Length 1092;
  Best Local Similarity  56.1%;  Pred. No. 1.9e-44;
  Matches  474;  Conservative    0;  Mismatches  359;  Indels   12;  Gaps    4;

Qy    184 AACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCC 243
Db    199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy    244 TGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAAC 303
Db    259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy    304 AAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCT 363
Db    319 GAGGCAGACGCCATTATGTTCCACCACCGGGACATT---TCTGATTTCAGTGACCTTCCT 375

Qy    364 CAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCAC 423
Db    376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy    424 ACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGAT 483
Db    436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy    484 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAA 543
Db    496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy    544 GTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 603
Db    553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy    604 AGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCA 663
Db    613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy    664 TTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTAT 723
Db    670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy    724 CTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCA 780
Db    730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
Db    790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
Db    850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
Db    910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
Db    970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy   1021 GTAAA 1025
Db   1030 GTCAA 1034


RESULT 15
GGU73678
LOCUS       GGU73678                2657 bp    DNA     linear   VRT 20-DEC-1996
DEFINITION  Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete
            cds.
ACCESSION   U73678
VERSION     U73678.1  GI:1657998
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N.,
            Schultz,T.L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
  TITLE     Molecular cloning and characterization of CFT1, a developmentally
            regulated avian alpha(1,3)-fucosyltransferase gene
  JOURNAL   J. Biol. Chem. 271 (51), 32960-32967 (1996)
  MEDLINE   97115837
REFERENCE   2  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M. and Larsen,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical
            Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES             Location/Qualifiers
     source          1. .2657
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            170. .1240
                     /gene="CFT1"
     CDS             170. .1240
                     /gene="CFT1"
                     /note="similar to mammalian fucosyltransferase FucTIV"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase"
                     /protein_id="AAC60060.1"
                     /db_xref="GI:1657999"
                     /translation="MELGPRWSPAARPGCPRRWRRRWALLGALLGAALALYVCVRELR
                     RRGSAAGRPEGEVTVLLWWEPFGRPWRPADCRRRYNITGCLLSADRGRYGEARAVLFH
                     HRDLALHGRQGLPRGPPPRPPRQRWVWMNFESPSHSPGLRGLAGLFNWTMSYRRDSDV
                     FVPYGYLYEPPSPRPFVLPRKSRLVAWVISNWNEEHARVRYYRQLKEHLPIDVYGARG
                     MALLEGSVVKTVSAYKFYLAFENSQHTDYITEKLWKNAFAASAVPVVLGPRRANYERF
                     IPADSFIHVDDFPSPRLLATYLKFLDKNKPSYRRYFAWRNKYEVHVTSFWDEHYCKVC
                     EAVRTAGNQLKTVQNLAGWFES"
BASE COUNT      581 a    712 c    818 g    546 t
ORIGIN
```

```
  Query Match             16.2%;  Score 175.2;  DB 5;  Length 2657;
  Best Local Similarity   52.9%;  Pred. No. 4.9e-36;
  Matches  453;  Conservative    0;  Mismatches  388;  Indels   15;  Gaps    3;

Qy     186 CGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTG 245
           |||     ||  | ||| | |||      ||| || || |   |||    ||   |   |||
Db     334 CGAGGTGACGGTGCTGCTGTGGTGGGAGCCCTTCGGCCGCCCCTGGCGCCCGGCCGACTG 393

Qy     246 CCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAA 305
           ||        | ||| |||   || ||||  || |   | ||||||     ||||  | |
Db     394 CCGGAGGCGCTACAACATCACGGGCTGCCTCCTGAGCGCCGACCGCGGGCGGTACGGCGA 453

Qy     306 ATCCCATGCGGTCCTGATTCACCATAGAGACAT---------CAGCTGGGATCTGACTAA 356
             | |  || || ||  | |||||  | ||| |          | || |  |   |
Db     454 GGCTCGGGCCGTGCTCTTCCACCACCGCGACCTGGCGCTGCACGGCCGCCAGGGGCTGCC 513

Qy     357 CTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACC 416
           |       |  | | |  | || ||    |||   ||| | |||||||| || ||||| ||
Db     514 CCGCGGGCCCCCGCCGCGACCCCCGCGGCAGCGGTGGGTGTGGATGAACTTCGAGTCGCC 573

Qy     417 CACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCG 476
           | | ||| |||||      | | ||| | |       | ||||||  |||  |  | || ||
Db     574 CTCGCACTCCCCCGGGCTGCGGGGCCTGGCCGGGCTCTTCAACTGGACCATGTCGTACCG 633

Qy     477 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGT 536
           || || || ||  ||   ||||| || || | | |    | |    |    ||
Db     634 ACGGGACTCGGACGTCTTCGTGCCCTACGGGTACCTCTACGAGCCGCCGTCGCCGCGGCC 693

Qy     537 GTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGA 596
           || |   ||||  ||||    | |||||  ||||||  | || |||||||||   ||
Db     694 CTTCGTGCTGCCTCGCAAAAGCCGGCTGGTGGCCTGGGTCATCAGCAACTGGAACGAGGA 753

Qy     597 GCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGG 656
           ||| ||| | ||    || |||    |||| |   ||   |   ||| ||   || ||
Db     754 GCACGCCCGCGTGCGCTACTACCGGCAGCTGAAGGAGCACCTGCCCATCGACGTGTACGG 813

Qy     657 CCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAA 716
             ||   |||         ||   ||     |   || |    ||  | ||  | |  ||
Db     814 ---GGCGCGCGGGATGGCGCTGCTGGAGGGCAGCGTGGTGAAGACGGTGTCGGCCTACAA 870

Qy     717 ATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT---A 773
           || || ||  | || || |||||    ||||  || |||||||| || ||||| |   |
Db     871 GTTCTACCTGGCCTTCGAGAACTCCCAGCACACGGACTACATCACCGAGAAGCTGTGGAA 930

Qy     774 CAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGA 833
           || || ||    ||  |  | || ||||||||||| || ||    ||||  ||||| ||
Db     931 GAACGCCTTCGCCGCCAGCGCGGTGCCTGTTGTCCTCGGCCCCCGCAGGGCCAACTACGA 990

Qy     834 GAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTT 893
           |   |  || || || || || ||||| || ||||| || ||   |   ||  |   | |
Db     991 GCGCTTCATCCCCGCCGACTCCTTCATCCACGTGGACGACTTCCCCAGCCCACGGCTGCT 1050

Qy     894 AGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTG 953
           || |  || ||||||    | || ||||||||       |||   || || ||   |||
Db    1051 GGCCACCTACCTGAAGTTCCTCGATAAAAACAAGCCCAGCTACAGGAGGTATTTCGCCTG 1110

Qy     954 GAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATG 1013
           | | ||  | | |   ||  || |  |    || |||||    |||   |||  ||  ||
Db    1111 GCGGAACAAGTATGAAGTCCACGTCACGTCTTTCTGGGATGAGCATTACTGCAAGGTTTG 1170

Qy    1014 CGACCATGTAAAAAGG 1029
           |||     || |  | |
Db    1171 CGAGGCCGTGAGGACG 1186


Search completed: October  6, 2002, 04:16:36
Job time : 1570.36 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 121.198 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-3_COPY_115_1194
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1080	100.0	2036	21	AAZ92672	Murine alpha-1,3-f
2	1080	100.0	2170	21	AAZ92645	Murine alpha-1,3-f
3	931.2	86.2	2676	21	AAZ92647	Human alpha-1,3-fu
4	931.2	86.2	2822	21	AAZ92646	Human alpha-1,3-fu
5	129	11.9	1814	18	AAT59505	Murine myeloid-lin
6	128.6	11.9	3594	18	AAT85219	Mouse alpha-fucosy
7	128.2	11.9	1086	15	AAQ56912	Lewis blood group
8	128.2	11.9	2042	18	AAT80111	Fucosyltransferase
9	128.2	11.9	2043	12	AAQ13330	Glycosyltransferas
10	128.2	11.9	2043	15	AAQ56906	DNA encoding a gly
11	128.2	11.9	2043	18	AAT76769	Human alpha 1,3/4
12	128.2	11.9	2043	18	AAT61675	Human alpha(1,3/1,
13	126.6	11.7	1316	15	AAQ56911	DNA encoding a gly
14	126.6	11.7	1316	18	AAT61679	Human alpha(1,3)-f
15	124.6	11.5	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
16	124.6	11.5	1654	18	AAT61680	Human alpha(1,3)-f
17	123	11.4	1126	21	AAF21135	Human low adenosin
18	123	11.4	1126	21	AAA35013	Human adenosine re
19	123	11.4	6944	21	AAF21137	Human low adenosin
20	123	11.4	6944	21	AAA35015	Human adenosine re
21	122.8	11.4	1701	15	AAQ77732	Human alpha-1,3-fu
22	122.8	11.4	1701	21	AAF21136	Human low adenosin
23	122.8	11.4	1701	21	AAA35014	Human adenosine re
24	120	11.1	1256	21	AAF21134	Human low adenosin
25	120	11.1	1256	21	AAA35012	Human adenosine re
26	120	11.1	1400	13	AAQ31436	Encodes a HeLa cel
27	120	11.1	1488	15	AAQ56910	DNA encoding a gly
28	120	11.1	2175	12	AAQ14382	Clone 7.2 encoding
29	120	11.1	2175	24	AAS17082	Human cDNA clone 7
30	120	11.1	2861	12	AAQ14383	Clone 1 encoding 1
31	120	11.1	2861	21	AAF21133	Human low adenosin
32	120	11.1	2861	21	AAA35011	Human adenosine re
33	120	11.1	2861	24	AAS17083	Human cDNA clone 1
34	120	11.1	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
35	120	11.1	3647	15	AAQ56909	DNA encoding a gly
36	120	11.1	3647	18	AAT61678	Human alpha(1,3)-f
37	118.4	11.0	2175	11	AAQ06691	CDX, a MILA involv
38	116.8	10.8	2134	18	AAT59506	Human myeloid deri
39	116	10.7	795	17	AAT13798	Alpha-(1,3/1/4)-fu
40	64.6	6.0	1578	23	ABL11867	Drosophila melanog
c 41	64.6	6.0	4229	23	ABL11866	Drosophila melanog
42	53	4.9	1353	21	AAZ51685	C. elegans alpha-1
43	43.6	4.0	2198	21	AAA97937	Mung bean alpha 1,
44	42.4	3.9	14041	22	AAH48024	Internal control B
45	40.8	3.8	398	22	AAF64428	Novel human polynu

ALIGNMENTS

RESULT 1
AAZ92672
ID AAZ92672 standard; cDNA; 2036 BP.
XX
AC AAZ92672;
XX
DT 05-JUN-2000 (first entry)
XX
DE Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; murine; mouse; ss.
XX
OS Mus sp.
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x

```
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Example 8; Page 162-164; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

  Query Match             100.0%;  Score 1080;  DB 21;  Length 2036;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-310;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 68

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 128

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 188

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 248

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 308

Qy     301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 368

Qy     361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 428

Qy     421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 488

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 548

Qy     541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 608

Qy     601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 668

Qy     661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 728

Qy     721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     729 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 788

Qy     781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     789 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 848

Qy     841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     849 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 908

Qy     901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     909 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 968

Qy     961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     969 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1028

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1029 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1088


RESULT 2
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag= a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
```

```
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match             100.0%;  Score 1080;  DB 21;  Length 2170;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-310;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 174

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 234

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 294

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 354

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 414

Qy     301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 474

Qy     361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 534

Qy     421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 594

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 654

Qy     541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 714

Qy     601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 774

Qy     661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 834

Qy     721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 894

Qy     781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 954

Qy     841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 1014

Qy     901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074

Qy     961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1134

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194

RESULT 3
AAZ92647
ID   AAZ92647 standard; cDNA; 2676 BP.
XX
AC   AAZ92647;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             295..1374
FT                   /*tag=  a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
```

```
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match             86.2%;  Score 931.2;  DB 21;  Length 2676;
  Best Local Similarity   91.4%;  Pred. No. 4.3e-266;
  Matches  987;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          |||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db     295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db     355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||| || || || ||||||||||||||||||| ||||| ||||| ||||||||||||
Db     415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db     475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||||||||| |||||||| |||||||||||||| ||||| ||| ||||||
Db     535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594

Qy     301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||| || ||||| || ||  |||||||||| |||||||||| || |||
Db     595 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 654

Qy     361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||| || |||||||||| |||||||||||||||||||||||| || ||||| |||
Db     655 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 714

Qy     421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||| |||||||||||||||||| ||||||| |||||||||| ||||| ||||||
Db     715 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 774

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          |||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db     775 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834

Qy     541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894

Qy     601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954

Qy     661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||| |||||||| || || ||||||||| ||||||| ||||||||| ||||||||||
Db     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014

Qy     721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          |||||||| |||||||| ||||| ||||| |||||||||| |||||||| ||||||||
Db    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 1074

Qy     781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          ||| ||||||| || |||||||||| ||||| |||||||||||||||||||||||||||
Db    1075 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1134

Qy     841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||| |||||||||||||||||||||| |||||||||||||| |||||||
Db    1135 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1194

Qy     901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||||||||||| |||||||||||||||| |||||||||||||||||||||| |||
Db    1195 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Qy     961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          ||||| |||||||| || ||||| ||||||||||||||||||  |||| ||||| |||
Db    1255 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1314

Qy    1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1374


RESULT 4
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag= a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
```

XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
DR P-PSDB; AAY80996.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Claim 4; Page 134-142; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC AAZ92646 and AAZ92647 represent cDNAs encoding human
CC alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

Query Match 86.2%; Score 931.2; DB 21; Length 2822;
Best Local Similarity 91.4%; Pred. No. 4.4e-266;
Matches 987; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db   1289 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 1348

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db   1349 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 1408

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db   1409 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 1468

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db   1469 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 1528

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db   1529 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 1588

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 360
          |||||||||||||| || ||||| || ||  |||||||||| |||||||||| || |||
Db   1589 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 1648

Qy    361 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 420
          |||||||| || |||||||||| |||||||||||||||||||||| || ||||| |||
Db   1649 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 1708

Qy    421 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 480
          ||||| |||||||||||||||||||| |||||||| |||||||||| ||||| ||||||
Db   1709 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 1768

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 540
          |||||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||
Db   1769 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| |||| || |||
Db   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          ||||| |||||||| || || ||||||||| ||||||| |||||||| ||||||||||
Db   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 780
          |||||||| |||||||| ||||| ||||| |||||||||| |||||||| ||||||||
Db   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 2068

Qy    781 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 840
          ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||||||||||
Db   2069 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 2128

Qy    841 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 900
          |||||||| |||||||||||||||||||||| |||||||||||||||| |||||||
Db   2129 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 2188

Qy    901 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||||||||||| |||||||||||||||| ||||||||||||||||||||||| |||
Db   2189 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248

Qy    961 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1020
          ||||| |||||||| || ||||| |||||||||||||||||  |||| ||||| |||
Db   2249 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 2308

Qy   1021 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2309 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 2368
```

RESULT 5
AAT59505
ID AAT59505 standard; cDNA; 1814 BP.
XX
AC AAT59505;
XX
DT 06-MAY-1997 (first entry)
XX
DE Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW septic shock; septicaemia; therapy; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 325..1353
FT /*tag= a
XX
PN WO9640881-A1.
XX
PD 19-DEC-1996.

```
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             11.9%;  Score 129;  DB 18;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 9.2e-28;
  Matches  335;  Conservative    0;  Mismatches  300;  Indels    6;  Gaps    2;

Qy     269 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 328
           | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     329 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 388
           |  | ||  |          |   ||  | |||| | | |    |||||||  |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     389 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 448
             ||| | |||       | || || |||| | | ||||          | || |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     449 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           | | ||||||  |   || |  ||||| ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     509 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 568
           ||||  |     | |    |||      |    | | ||   |||    | | |||
Db     788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 628
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     629 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 688
           |     | |  |  |  |    |  || |  ||   ||||     | |   || |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     689 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 748
           ||||  | |||||  |  |    |      || || ||  | ||||| ||||||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     749 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 805
             || |||||||| || ||| |||    ||||||  |   ||||||| | ||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     806 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 865
           |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     866 TGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 906
           . |||| || || | |||| || ||||  | ||    |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag=  a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag=  b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag=  c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag=  d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag=  e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag=  f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag=  g
FT                   /number=  3
FT   exon            2067..3552
FT                   /*tag=  f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
```

```
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             11.9%;  Score 128.6;  DB 18;  Length 3594;
  Best Local Similarity   52.5%;  Pred. No. 1.7e-27;
  Matches  331;  Conservative    0;  Mismatches  294;  Indels    6;  Gaps    2;

Qy     269 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 328
           | |||| ||| |   |  ||||     ||      ||   |  |||| ||  |  | ||||
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336

Qy     329 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 388
           |  | ||  |           |   ||  | |||| | | |    |||||| |   |||
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     389 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 448
             ||| | |||       | || || |||| | | ||||          | || |
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     449 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           | | ||||||  |   || |  ||||| ||||||||||||||||   || || || ||
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     509 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 568
           ||||  |     | |    |||       |    | | ||  |||    | | |||
Db    2517 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 628
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     629 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 688
           |     | |  |  | |     |  || |  ||   ||||     | |   || |||
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     689 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 748
           |||| | |||||  |  |     |     || || ||  | ||||| ||||||   ||
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy     749 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 805
           ||  |||||||| || ||| |||    ||||||  |   |||||| | ||||| || |
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy     806 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 865
           |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy     866 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 896
           |||| || || | |||| || ||||  | ||
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904
```

```
RESULT 7
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
DE   Lewis blood group fucosyltransferase DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding a Lewis blood group fucosyl
CC   transferase (Fuc-TIII).  The enzyme produced by the DNA may be
CC   non glycosylated.  This prevents premature loss of enzyme activity.
CC   It can also be used in in vitro reactions to modify cell surface
CC   oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-11.
XX
SQ   Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

  Query Match             11.9%;  Score 128.2;  DB 15;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 1.2e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db     271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |   |   |  | | |  | | || |       | ||||| |     |||
Db     331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
             ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | | ||
Db     391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
           |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |     |    ||          | |  |  |  |   ||||
Db     511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
           || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |   ||         | |  |  |  ||  || || |  |  |  |   | | ||
```

```
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 8
AAT80111
ID   AAT80111 standard; cDNA; 2042 BP.
XX
AC   AAT80111;
XX
DT   01-DEC-1997  (first entry)
XX
DE   Fucosyltransferase III coding sequence.
XX
KW   P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW   calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW   thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW   inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW   disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW   adult respiratory distress syndrome; tumour metastasis;
KW   fucosyltransferase III; acetylglucosaminyltransferase;
KW   autoimmune disorder; inflammatory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9706176-A2.
XX
PD   20-FEB-1997.
XX
PF   02-AUG-1996;   96WO-US12820.
XX
PR   17-MAY-1996;   96US-0649802.
PR   03-AUG-1995;   95US-0510920.
PR   15-MAY-1996;   96US-0017794.
XX
PA   (OKLA ) UNIV OKLAHOMA STATE.
XX
PI   Cummings RD,  Mcever RP,  Moore KL;
XX
DR   WPI; 1997-154206/14.
XX
PT   New O-glycan(s) derived from P-selectin glyco:protein ligand
PT   (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT   for treating inflammatory or autoimmune disorders or tumours
XX
PS   Disclosure; Page 80-81; 96pp; English.
XX
CC   This sequence represents the coding sequence for the human
CC   fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC   acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC   expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC   mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC   protein expressed on the surfaces of activated platelets and endothelium
CC   in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC   high affinity P-selectin ligand produced by leukocytes. Binding of
CC   P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC   treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC   highly extended extracellular domain, which is a feature of mucin-like
CC   proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC   the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC   O-glycans, and O-glycanated products of the invention. The products are
CC   used for inhibiting the binding of P-selectin and other selectins to
CC   PSGL-1. They can be used for the treatment of inflammation, injury
CC   resulting from ischaemia and reperfusion, bacterial sepsis and
CC   disseminated intravascular coagulation, adult respiratory distress
CC   syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC   other autoimmune or inflammatory disorders. The products can also be used
CC   in the production of antibodies and in diagnostic applications.
XX
SQ   Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

  Query Match             11.9%;  Score 128.2;  DB 18;  Length 2042;
  Best Local Similarity   53.0%;  Pred. No. 1.7e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330

Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 9
AAQ13330
ID   AAQ13330 standard; DNA; 2043 BP.
```

```
XX
AC   AAQ13330;
XX
DT   07-NOV-1991  (first entry)
XX
DE   Glycosyltransferase gene in clone pCDM7-alpha(1,3/1,4)FT.
XX
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..1158
FT                   /*tag=  a
FT   polyA_site      2023..2034
FT                   /*tag=  b
FT   polyA_signal    1926..2003
FT                   /*tag=  c
FT   repeat_region   1183..1547
FT                   /*tag=  d
FT                   /note= "20 copies of a degenerate 16bp sequence"
FT   repeat_region   1885..1999
FT                   /*tag=  e
FT                   /note= "similar to human Alu sequences"
XX
PN   WO9112340-A.
XX
PD   22-AUG-1991.
XX
PF   14-FEB-1991;   91WO-US00899.
XX
PR   12-DEC-1990;   90US-0627621.
PR   14-FEB-1990;   90US-0479858.
PR   14-FEB-1990;   90US-0480133.
XX
PA   (UNMI ) UNIV OF MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1991-267151/36.
DR   P-PSDB; AAR13749.
XX
PT   Isolation of gene conveying post-translational characteristic -
PT   e.g. the presence of soluble or membrane bound oligo or
PT   polysaccharide or glycosyltransferase.
XX
PS   Claim 11; Fig 1; 155pp; English.
XX
CC   The DNA encodes a protein sequence capable of functioning as a
CC   GDP- Fuc:[beta-D-Gal(1,4/1,3)]-D-GlcNAc(/Glc)alpha (1,3/1,4)-
CC   fucosyltransferase. The sequence coded by nucleotides 127 to 1083
CC   represents the functional protein. The DNA sequence can be used to
CC   construct animal cell lines with respect to posttranslational
CC   modification of the oligosaccharides on cell-surface, intracellular
CC   or secreted proteins or lipids by sub-terminal alpha(1,3) and alpha
CC   (1,4) fucose residues that represent the products of this enzyme.
CC   These can be used for the prodn. of diagnostics and therapeutics.
CC   See also AAQ13331-Q13333.
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match             11.9%;  Score 128.2;  DB 12;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.7e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           |||||  ||||  | ||||||    |||||     |  |  |  |||||  |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |   |   |  | |   | | || |      | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
           |||||  ||| | ||  ||  |||  |||||  ||  ||  ||  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
           |  |||||  ||  ||  |  | ||  |||  | ||  ||  ||  |||     |||  ||  |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |     |    ||          | |   |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
           || |||||  |||||  |||  ||||||||  || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
           | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
           ||  | |||   || ||||||||| || |||||  |     ||  ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
           || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 10
AAQ56906
ID   AAQ56906 standard; DNA; 2043 BP.
XX
AC   AAQ56906;
XX
DT   26-JUL-1994  (first entry)
XX
DE   DNA encoding a glycosyltransferase.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             72..1158
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
```

```
DR   P-PSDB; AAR45934.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 1; 249pp; English.
XX
CC   The sequence is that encoding human glycosyl transferase.  The enzyme
CC   produced by the DNA may be non glycosylated.  This prevents premature
CC   loss of enzyme activity.  It can also be used in in vitro reactions to
CC   modify cell surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-12.
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match            11.9%;  Score 128.2;  DB 15;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.7e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |   |   |  | |   | | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
           |||||  ||| | || || ||| |||||  || ||  ||  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             | ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |    |   ||          | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
            || |||||  |||||  |||   |||||||| || ||    ||||||||     || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |    ||          | |  |  |   ||  || || |   |   |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
             | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
           ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
           || || || || ||||| ||    ||    ||||| |||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||  || |      ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 11
AAT76769
ID   AAT76769 standard; cDNA; 2043 BP.
XX
AC   AAT76769;
XX
DT   15-SEP-1997  (first entry)
XX
DE   Human alpha 1,3/4 fucosyltransferase cDNA.
XX
KW   GDP-L-fucose:beta-D-N-acetylglucosamide 3/4-alpha-L-fucosyltransferase;
KW   human; alpha 1,2 FT; transgene; transgenic mouse; animal model;
KW   intestinal adhesion; Helicobacter pylori infection; stomach;
KW   small intestine; gut; epithelial cell; surface receptor;
KW   carbohydrate antigen; gastritis; peptic ulcer; neoplasia;
KW   gastric adenocarcinoma; Lewis antigen; fucosylation; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             104..1201
FT                   /*tag= a
XX
PN   US5625124-A.
XX
PD   29-APR-1997.
XX
PF   11-JUL-1994;   94US-0273411.
XX
PR   11-JUL-1994;   94US-0273411.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Falk P,  Gordon JI;
XX
DR   WPI; 1997-258275/23.
DR   P-PSDB; AAW23806.
XX
PT   Animal model for Helicobacter pylori infection - comprising
PT   transgenic mouse expressing human enzyme promoting intestinal
PT   adhesion
XX
PS   Example 1; Columns 23-26; 24pp; English.
XX
CC   A claimed transgenic mouse expresses, in its intestinal epithelial
CC   cells, the enzyme human GDP-L-fucose:beta-D-galactoside 2-alpha-L-
CC   fucosyltransferase (also called alpha 1,2 FT) or human GDP-L-fucose:
CC   beta-D-N-acetylglucosamide 3/4-alpha-L-fucosyltransferase (also
CC   called alpha 1,3/4 FT). The enzyme is expressed under the
CC   control of a gut epithelial cell-specific promoter and Helicobacter
CC   pylori adheres to the transgenic cells. The transgenic mouse and
CC   intestinal epithelial cells from it are useful as models for screening
CC   compounds for the ability to inhibit adhesion of H. pylori to gut
CC   epithelial cells. The present sequence encodes human alpha 1,3/4 FT
CC   and was published by Kukowska-Latallo et al., Genes Dev., 4:
CC   1288-1303 (1990).
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match            11.9%;  Score 128.2;  DB 18;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.7e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |   |   |  | |   | | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
           |||||  ||| | || || ||| |||||  || ||  ||  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             | ||||| || ||  |  | || ||| | || || || |||     ||| || |||
```

```
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 12
AAT61675
ID   AAT61675 standard; cDNA; 2043 BP.
XX
AC   AAT61675;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3/1,4)-fucosyltransferase (Lewis enzyme) cDNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TIII;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..1158
FT                   /*tag= a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13638.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 3; Page 268-269; 329pp; English.
XX
CC   A cDNA clone (AAT61675) codes for an enzyme (AAW13638) that functions as
CC   a GDP-Fuc:(beta-D-Gal(1,4/1,3))-D-GlcNAc(/Glc) alpha(1,3/1,4)-
CC   fucosyltransferase (Fuc-TIII) which, when expressed by the cloned
CC   cDNA sequence, functions within mammalian cells to generate de novo
CC   expression of specific SSEA-1 (Lewis x), sialyl-Lewis x, Lewis a and
CC   sialyl-Lewis a structures.  The clone was identified by transfecting
CC   COS-1 cells with genomic DNA from human A431 cells, and isolating a
CC   cloned cDNA that detd. expression of surface-localised SSEA-1
CC   structures.  The cDNA can be used to construct animal cell lines
CC   with specific capabilities with respect to post-translational
CC   modification of the oligosaccharides of expressed proteins or
CC   lipids, to produce recombinant Fuc-TIII for use in oligosaccharide
CC   prodn., to genotype individuals for the Lewis locus, and in the
CC   prodn. of chimeric fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 2043 BP; 413 A; 646 C; 576 G; 408 T; 0 other;

  Query Match             11.9%;  Score 128.2;  DB 18;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.7e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330

Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 13
AAQ56911
```

```
ID   AAQ56911 standard; DNA; 1316 BP.
XX
AC   AAQ56911;
XX
DT   26-JUL-1994  (first entry)
XX
DE   DNA encoding a glycosyltransferase.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45939.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding human glycosyl transferase.  The enzyme
CC   produced by the DNA may be non glycosylated.  This prevents premature
CC   loss of enzyme activity.  It can also be used in in vitro reactions to
CC   modify cell surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-12.
XX
SQ   Sequence 1316 BP; 236 A; 456 C; 372 G; 252 T; 0 other;

  Query Match             11.7%;  Score 126.6;  DB 15;  Length 1316;
  Best Local Similarity   52.9%;  Pred. No. 4e-27;
  Matches  373;  Conservative    0;  Mismatches  314;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
          ||| |  ||||  | |||  |    |||||    |  |  | |||||| | | ||||| 
Db    430 TGCAACATCACTGCCGACTCCAGTGTGTACCCACAGGCAGACGCGGTCATCGTGCACCAC 489

Qy    331 AGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
            | || ||||  |   | |  |   | ||| | || |  |   |||||||  |     |||
Db    490 TGGGATATCATGTACAACCCCAGTGCCAACCTCCCGCCCCCCACCAGGCCGCAGGGGCAG 549

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
          |||||| ||| | |   | ||||| ||||   ||  ||   ||   | | | ||
Db    550 CGCTGGATCTGGTTCAGCATGGAGTCCCCCAGCAACTGCCGGCACCTGGAAGCCCTGGAC 609

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    610 GGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 669

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
          |   ||  |  | |  |    |   ||           | |  |  |  |   ||||
Db    670 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 729

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
          || |||||  |||||  |||   |||||||| || ||    ||||||||     || |||
Db    730 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCGGCCAGGGTGCGCTACTAC 789

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
          |   ||         | |  | |  ||   || || |   |   |   |   | ||
Db    790 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 846

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
          | |   |  ||||   ||  | ||    |  || || ||||| | || || |||
Db    847 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTATCTGGCCTTCGAGAAC 906

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    907 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 966

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
          || || || || ||||| ||    ||    ||||| |||| |   | || | ||  |
Db    967 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCGCCCGACGCC 1026

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
          ||||| || ||||| || ||  |     ||||  ||  | ||    || ||| |||| |
Db   1027 TTCATCCACGTGGATGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 1086

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          ||||| || |      ||||| || |||||  |||| |  ||
Db   1087 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1131


RESULT 14
AAT61679
ID   AAT61679 standard; DNA; 1316 BP.
XX
AC   AAT61679;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TV) DNA.
XX
KW   Alpha(1,3)-fucosyltransferase; Lewis enzyme; Fuc-TV;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             121..1245
FT                   /*tag= a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13642.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 5; Page 287-288; 329pp; English.
XX
CC   A DNA clone (AAT61679) codes for an enzyme (AAW13642) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TV) which can utilise neutral type II oligosaccharide
CC   precursors but which cannot efficiently utilise type I
CC   glycococonjugates.  It was identified in a human lambda phage
CC   genomic DNA library using a probe obtd. from Lewis enzyme cDNA (see
CC   also AAT61675).  The Fuc-TV enzyme, when expressed by the cloned DNA
CC   sequence, functions within mammalian cells to generate de novo
CC   expression of specific cell surface glycoconjugate structures that
CC   are recognised by antibodies against SSEA-1, Lewis x or sialyl
CC   Lewis x.
XX
```

```
SQ   Sequence 1316 BP; 236 A; 456 C; 372 G; 252 T; 0 other;

  Query Match           11.7%;  Score 126.6;  DB 18;  Length 1316;
  Best Local Similarity  52.9%;  Pred. No. 4e-27;
  Matches  373;  Conservative    0;  Mismatches  314;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
          ||| |  ||||  | |||  |    |||||    |  |  | ||||||  |  | |||||
Db    430 TGCAACATCACTGCCGACTCCAGTGTGTACCCACAGGCAGACGCGGTCATCGTGCACCAC 489

Qy    331 AGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |   | |  |   | ||| | || |  |   |||||||  |     |||
Db    490 TGGGATATCATGTACAACCCCAGTGCCAACCTCCCGCCCCCCACCAGGCCGCAGGGGCAG 549

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
            ||||| ||| | |   | ||||| ||||   ||  ||  ||   |   | | | ||
Db    550 CGCTGGATCTGGTTCAGCATGGAGTCCCCCAGCAACTGCCGGCACCTGGAAGCCCTGGAC 609

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
           |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    610 GGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 669

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
          |   ||  |  | |  |     |    ||          | |  |  |  |   ||||
Db    670 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 729

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
           || |||||  |||||  |||  |||||||| || ||    ||||||||     || |||
Db    730 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCGGCCAGGGTGCGCTACTAC 789

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |    ||          | |  |  |  ||  || || |   |   |   |   |  | ||
Db    790 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 846

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
           | |    |  ||||    ||  | ||    |  || || |||||· | || || |||
Db    847 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTATCTGGCCTTCGAGAAC 906

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    907 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 966

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
          || || || || ||||| ||    ||   |||||  |||| |   |  || |  | ||  |
Db    967 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCGCCCGACGCC 1026

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db   1027 TTCATCCACGTGGATGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 1086

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1087 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1131


RESULT 15
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match           11.5%;  Score 124.6;  DB 15;  Length 1654;
  Best Local Similarity  52.1%;  Pred. No. 1.8e-26;
  Matches  408;  Conservative    0;  Mismatches  354;  Indels   21;  Gaps    5;

Qy    196 ATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
          || ||| | |||   |||||  |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    256 TTCAAT---ATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 312
          |   |   |      |  ||| |  ||||  | ||||||    ||||     |  | |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    313 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAG 369
          |||||| |  | |||||  ||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    370 GCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCC 429
          | ||||  |     |||  ||||| ||| | |   |  ||||| || |  |||  |
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    430 CAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGAT 489
          ||   ||  | ||| ||    |  ||||| || ||  |  | || ||| | || ||  |
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    490 ATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTT 540
          |||     ||| ||  ||||   ||  |  | |  |     |    ||          | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    541 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |  |  |   ||||    || ||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    601 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 660
          ||||||||    || ||| |    ||          | |  |  |  ||   || ||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    661 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 720
          | |    |  | ||  | |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    721 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTAC---AAT 777
          |||||  | || || |||||  | |||   || ||||||||| || ||||| |     ||
```

```
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    778 GCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAAT 837
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    838 TATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCA 897
          |   | ||| | ||  | ||||| || ||||| || ||  |   |||| ||  | ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    898 AAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGA 957
             || ||| ||||  | |||||  || |       ||||| || |||||  |||| |
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    958 AAG 960
            ||
Db   1095 GAG 1097
```

Search completed: October 5, 2002, 23:10:19
Job time : 136.198 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 904.255 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-3_COPY_115_1194
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	661.8	61.3	697	9	BB397507	BB397507 BB397507
2	580.8	53.8	607	10	BI102760	BI102760 602888338
3	368	34.1	649	9	AV338075	AV338075 AV338075
4	366	33.9	716	9	AV725071	AV725071 AV725071
5	335.8	31.1	648	9	BB595164	BB595164 BB595164
6	315	29.2	625	9	BB649967	BB649967 BB649967
c 7	304.8	28.2	1101	12	CNS059ST	AL327638 Tetraodon
8	303.8	28.1	767	9	BB640911	BB640911 BB640911
c 9	272.2	25.2	579	10	BM021081	BM021081 ie75f03.x
10	255.8	23.7	628	9	BB651409	BB651409 BB651409
11	209.4	19.4	568	10	BM021317	BM021317 ie75f03.y
c 12	179.6	16.6	1100	12	CNS0579Z	AL324368 Tetraodon
13	175.8	16.3	844	12	CNS01V8V	AL168808 Tetraodon
c 14	171.4	15.9	570	10	BM082484	BM082484 fu25e08.x
c 15	164.8	15.3	887	12	CNS02Q0S	AL208693 Tetraodon
16	157.4	14.6	670	9	BB626126	BB626126 BB626126
c 17	156.8	14.5	1016	12	CNS0429P	AL271222 Tetraodon
18	152.2	14.1	199	9	BB398103	BB398103 BB398103
c 19	148.4	13.7	666	10	BJ050300	BJ050300 BJ050300
20	145.8	13.5	651	10	BJ036855	BJ036855 BJ036855
21	142	13.1	640	10	BJ073227	BJ073227 BJ073227
c 22	140.4	13.0	711	12	CNS034XF	AL228012 Tetraodon
c 23	133.4	12.4	501	12	B41073	B41073 HS-1052-B2-
c 24	118.6	11.0	674	10	BE605503	BE605503 fl10h04.x
25	115.6	10.7	971	12	CNS047IV	AL278032 Tetraodon
26	115	10.6	533	10	BI604066	BI604066 603243886
c 27	110.8	10.3	1023	12	CNS02ANG	AL188773 Tetraodon
c 28	110.2	10.2	615	10	BI326811	BI326811 STR00078
29	106.2	9.8	487	10	BE863396	BE863396 UI-M-BH0-
30	106.2	9.8	720	10	BF180380	BF180380 601804564
c 31	105.8	9.8	702	10	BJ043578	BJ043578 BJ043578
32	90	8.3	863	12	CNS02ROF	AL210840 Tetraodon
33	88.4	8.2	551	10	BJ035701	BJ035701 BJ035701
c 34	87.8	8.1	566	10	BJ090565	BJ090565 BJ090565
35	85.4	7.9	601	10	BJ036679	BJ036679 BJ036679
36	85.4	7.9	605	10	BJ028804	BJ028804 BJ028804
c 37	84.8	7.9	776	12	CNS03RIO	AL257289 Tetraodon
c 38	83.8	7.8	1056	12	CNS01VL7	AL169252 Tetraodon
39	82.6	7.6	638	10	BJ035697	BJ035697 BJ035697
40	81.8	7.6	655	10	BG538740	BG538740 602566977
c 41	81.4	7.5	878	12	CNS02V1D	AL215194 Tetraodon
42	80.8	7.5	304	9	AU081478	AU081478 AU081478
43	80.8	7.5	1057	12	CNS03ADC	AL235065 Tetraodon
44	74.8	6.9	838	10	BG432556	BG432556 602500710
45	74.6	6.9	590	10	BM082755	BM082755 fu25e08.y

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN

  Query Match             61.3%;  Score 661.8;  DB 9;  Length 697;
  Best Local Similarity   99.6%;  Pred. No. 4.1e-175;
  Matches  674;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy     404 ATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGA 463
           | ||||||||||||||||||||||| |||||||||||| |||||||||||||||||||
Db       4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy     464 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy     524 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy     584 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 643
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy     644 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 703
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy     704 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 763
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy     764 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 823
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy     824 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 883
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy     884 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 943
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy     944 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 1003
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy    1004 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 1063
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy    1064 AATGGTTTTGGAATTAA 1080
           |||||||||||||||||
Db     663 AATGGTTTTGGAATTAA 679


RESULT 2
BI102760
LOCUS       BI102760                607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: l  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN

  Query Match             53.8%;  Score 580.8;  DB 10;  Length 607;
  Best Local Similarity   99.2%;  Pred. No. 2.4e-152;
  Matches  604;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;
```

```
Qy    223 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 282
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy    283 ACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 342
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy    343 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATG 402
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy    403 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy    463 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 522
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy    523 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy    583 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy    643 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy    703 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 762
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy    763 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGG 822
          |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db    539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy    823 GAAAACTAT 831
          |||||||||
Db    599 GAAAACTAT 607
```

RESULT 3
AV338075

```
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .649
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="6430407H24"
                      /clone_lib="RIKEN full-length enriched, adult male
                      olfactory bulb"
                      /sex="male"
                      /tissue_type="olfactory brain"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 100.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                      3']. cDNA was cloned into the XhoI and BamHI sites.
                      Vector: a modified pBluescript KS(+) after bulk excision
                      from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                      BamHI"
BASE COUNT      209 a    105 c    117 g    218 t
ORIGIN
```

```
  Query Match             34.1%;  Score 368;  DB 9;  Length 649;
  Best Local Similarity   98.7%;  Pred. No. 1.8e-92;
  Matches  371;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    705 ATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGA 764
```

```
Db      2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61
Qy    765 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 824
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121
Qy    825 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 884
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181
Qy    885 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 944
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241
Qy    945 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 1004
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301
Qy   1005 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 1064
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361
Qy   1065 ATGGTTTTGGAATTAA 1080
Db    362 ATGGTTTTGGAATTAA 377

RESULT 4
AV725071
LOCUS       AV725071                716 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION   AV725071
VERSION     AV725071.1  GI:10830091
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
FEATURES    This clone is available at CHGC in Shanghai.
     source          Location/Qualifiers
                     1. .716
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCBLB02"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      219 a      113 c      133 g      249 t        2 others
ORIGIN

Query Match          33.9%;  Score 366;  DB 9;  Length 716;
Best Local Similarity   91.7%;  Pred. No. 6.7e-92;
Matches  387;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    659 AAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAAT 718
Db      1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60
Qy    719 TTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATG 778
Db     61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120
Qy    779 CATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATT 838
Db    121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180
Qy    839 ATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAA 898
Db    181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240
Qy    899 AATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAA 958
Db    241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300
Qy    959 AGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACC 1018
Db    301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360
Qy   1019 ATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1078
Db    361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420
Qy   1079 AA 1080
Db    421 AA 422

RESULT 5
BB595164
LOCUS       BB595164                648 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB595164 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION   BB595164
VERSION     BB595164.2  GI:16450167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11491766.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T , Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M , Inoue,Y , Kira,A  and
            Hayashizaki,Y.
```

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Location/Qualifiers
source 1. .648
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030004N01"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 141 a 206 c 138 g 163 t
ORIGIN

Query Match 31.1%; Score 335.8; DB 9; Length 648;
Best Local Similarity 99.4%; Pred. No. 2e-83;
Matches 337; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 369

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 429

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 489

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 549

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    550 TCCTGCCAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 609

Qy    301 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATC 339
          ||||||||||||||||||||||||||||| |||||||
Db    610 AACAAATCCCATGCGGTCCTGATTCACCATATAGACATC 648
```

RESULT 6
BB649967
LOCUS BB649967 625 bp mRNA linear EST 26-OCT-2001
DEFINITION BB649967 RIKEN full-length enriched, 16 days embryo head Mus musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION BB649967
VERSION BB649967.1 GI:16484222
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 625)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A. , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .625
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C130099N17"
/clone_lib="RIKEN full-length enriched, 16 days embryo head"

```
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="16 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT       136 a    198 c    135 g    156 t
ORIGIN

  Query Match                    29.2%;  Score 315;  DB 9;  Length 625;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-77;
  Matches  315;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 370

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 430

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     431 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 490

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     491 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 550

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     551 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 610

Qy     301 AACAAATCCCATGCG 315
             |||||||||||||||
Db     611 AACAAATCCCATGCG 625


RESULT 7
CNS059ST/c
LOCUS       CNS059ST                1101 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            045P22 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL327638
VERSION     AL327638.1  GI:8221227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C ,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis usıng
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="045P22"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA045DH11A1-end : T3"
BASE COUNT       287 a    249 c    269 g    278 t     18 others
ORIGIN

  Query Match                    28.2%;  Score 304.8;  DB 12;  Length 1101;
  Best Local Similarity  58.7%;  Pred. No. 1.3e-74;
  Matches  608;  Conservative    0;  Mismatches  418;  Indels    9;  Gaps    5,

Qy      50 TCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGG 109
             ||||| |||| ||||     || |                 || ||    || ||  ||| ||||| |
Db    1064 TCATCGTGGGGTGCTNTGTGACTGTGTNNTTTATGTACTNTAATCCATCCACCAGCT-GN 1006

Qy     110 TCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAA 169
             | | | ||||  |||||  | |  |                 | | | || ||  ||| | || || |
Db    1005 TTTCCGGTCCTGTGGAGGCAGGACTGAGCTCGGAGCATGTCAAGAATCAACTATCAACCA 946

Qy     170 AAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCT 229
             | | |||| |     |||  || |  || ||| | ||| | ||||| || || || ||||
Db     945 AGAGTGATAAA---AACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCT 889

Qy     230 TTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACC 289
             | ||||| |     ||| | ||  |||  | | || | | | | | || || | |||
Db     888 TCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACA 829

Qy     290 GCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATC 349
             | |  | |||||  ||| ||  | | ||||| |||||||||  | | | |
Db     828 GAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACC 769

Qy     350 TGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAG 409
             || | ||| | ||     | || || |||||||||||| | | | | |
Db     768 TGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGG 709

Qy     410 AGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAA 469
             |||| ||| | ||| | |  |  | ||| || |  | |||||||||||||| || |
Db     708 AGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGA 649

Qy     470 CTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATC 529
             |||||  |||||  || ||     | |  |  | | | |  ||  |||  ||
Db     648 ACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATG 589

Qy     530 CCTTTGTG---TTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACT 586
             | | ||||   ||| |||||  | ||  |||| || ||| |||| ||||| ||||
Db     588 CTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACT 529

Qy     587 GGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCC 646
             |||||| || |||||  | || ||  |||||||| |  |||| | |||| |||
Db     528 GGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCC 469

Qy     647 ACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATAT 706
             |||| |||||  | |||||||||  || || |  |  ||| || |  | || |||||
Db     468 ACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCG 409
```

```
Qy     707 CTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAA 766
           | |  || || || || || |||||||| ||||| || |||||||| ||| | || ||||
Db     408 CCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAA 349

Qy     767 AGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAA 826
           | || |||||  || ||   |  ||  |||| || || || ||||| ||| || |||| |
Db     348 AACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGA 289

Qy     827 ACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCA 886
           ||||  |||||    || |  |  ||| | || || || |||||||| |||  |||  |||
Db     288 ACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCA 229

Qy     887 GTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACT 946
             ||| |  ||  | ||| |   |    | ||||||||  |  |  |||||||||  || |
Db     228 AGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATT 169

Qy     947 TTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCC 1006
           || |   |  |||   ||      ||    |  || || |          | || |
Db     168 TTCANGTGGCGAAGACACTTTAAAGCAA-GGGCAGTGGCATTTGCCCGAGGCCACCTGTC 110

Qy    1007 TGGCATGCGACCATGTAAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAA 1065
            ||| || ||  | || |  |||| | |  |||| | ||| |   |  |  ||| | ||
Db     109 AGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAG 50

Qy    1066 TGGTTTTGGAATTAA 1080
           ||||  |||  || |
Db      49 TGGTACTGGGGTTGA 35
```

RESULT 8
BB640911
LOCUS BB640911 767 bp mRNA linear EST 26-OCT-2001
DEFINITION BB640911 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730093G08 5', mRNA sequence.
ACCESSION BB640911
VERSION BB640911.1 GI:16476069
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 767)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .767
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730093G08"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 175 a 239 c 176 g 172 t 5 others
ORIGIN

Query Match 28.1%; Score 303.8; DB 9; Length 767;
Best Local Similarity 96.3%; Pred. No. 2.2e-74;
Matches 311; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy       1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     347 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 406

Qy      61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     407 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 466

Qy     121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     467 ATGGAGTCTGCAAGTTCTGGGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 526

Qy     181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 586

Qy     241 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 300
           ||| |||||||||||||||||| ||| ||||||| |||||| | || ||||||  ||
Db     587 TCCGGCCAAGCAATGTTCAATATCCAGGGGGGCCATCTTACAACAAAACGGTCATTGGAC 646

Qy     301 AACAAATCCCATGCGGTCCTGAT 323
           |||||||||| | || ||||||
Db     647 AACAAATCCCAGGGGGGCCTGAT 669
```

RESULT 9

```
BM021081/c
LOCUS       BM021081                579 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021081
VERSION     BM021081.1  GI:16535437
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES            Location/Qualifiers
     source         1. .579
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                    /sex="Both"
                    /tissue_type="Islets of Langerhans"
                    /dev_stage="Adult"
                    /lab_host="DH10B"
                    /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                    Site_2: Sal 1; Starting library constructed using
                    SuperScript Plasmid Library kit (Life Technologies). cDNA
                    made by oligo-dT priming. Size-selected by column
                    fractionation; average insert size 1.08 kb. Library was
                    amplified once on solid support and plasmid DNA from
                    library was prepared. The library DNA was normalized by
                    method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                    Research 6:791-806; 0.5 microgram single-stranded library
                    plasmid DNA was mixed with 5 micrograms PCR product
                    representing library inserts and hybridized to an Ecot of
                    20. Single-stranded (unhybridized) plasmids were isolated
                    by hydroxyapatite chromatography and used to make this
                    library."
BASE COUNT      203 a    103 c     93 g    180 t
ORIGIN
```

```
Query Match          25.2%;  Score 272.2;  DB 10;  Length 579;
Best Local Similarity  92.6%;  Pred. No. 1.5e-65;
Matches  286;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy    772 TACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTAT 831
          ||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy    832 GAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAG 891
          |||||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy    892 TTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAAC 951
```

```
          ||||||| ||||||||||||| |||||||||||||||||| ||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy    952 TGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCA 1011
          ||||| |||||||| |||||||| || |||| |||||||||||||||||||| ||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy   1012 TGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 1071
          ||||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy   1072 TGGAATTAA 1080
          |||||||||
Db    279 TGGAATTAA 271
```

```
RESULT 10
BB651409
LOCUS       BB651409                628 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB651409 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230079N02 5', mRNA sequence.
ACCESSION   BB651409
VERSION     BB651409.1  GI:16485662
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES    e mouse tissues.
                    Location/Qualifiers
```

```
     source          1. .628
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C230079N02"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 479.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      129 a    200 c    143 g    152 t      4 others
ORIGIN

  Query Match           23.7%;  Score 255.8;  DB 9;  Length 628;
  Best Local Similarity   96.3%;  Pred. No. 6.5e-61;
  Matches  260;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 406

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 466

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 526

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          |||||||||||||||||||| || ||||||||||| |||||||||| ||||||||||
Db    527 GTTAACGAAACTACCATTCTGGNTTAGGTATGGCCATNTGGGCAGACCTNTGACCTTACA 586

Qy    241 TCCTGCCAAGCAATGTTCAATATCCAAGGG 270
          |||||||||   |||||  ||||||||||||
Db    587 TCCTGCCAAAGCATGTTTCATATCCAAGGG 616


RESULT 11
BM021317
LOCUS       BM021317                 568 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021317
VERSION     BM021317.1  GI:16535673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ie75f03.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT      127 a    170 c    123 g    148 t
ORIGIN

  Query Match           19.4%;  Score 209.4;  DB 10;  Length 568;
  Best Local Similarity   91.4%;  Pred. No. 7.1e-48;
  Matches  222;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 60
          |||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db    326 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 385

Qy     61 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    386 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 445

Qy    121 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db    446 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 505

Qy    181 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db    506 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 565

Qy    241 TCC 243
          |||
Db    566 TCC 568


RESULT 12
CNS0579Z/c
LOCUS       CNS0579Z                 1100 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            007H05 of library A from Tetraodon nigroviridis, genomic survey
```

```
            sequence.
ACCESSION   AL324368
VERSION     AL324368.1  GI:9557250
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007H05"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA007CD03C1~end : T7"
BASE COUNT       314 a    228 c    252 g    304 t      2 others
ORIGIN
```

```
  Query Match                    16.6%;  Score 179.6;  DB 12;  Length 1100;
  Best Local Similarity          55.6%;  Pred. No. 2.1e-39;
  Matches  430;  Conservative      0;  Mismatches  331;  Indels   13;  Gaps    4;

Qy    195 CATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAAT 254
Db    774 CATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTTGCATACGACTCTTGTAAAGTTTA 715
Qy    255 GTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGC 314
Db    714 CTACAATATTGACGGCTGTGAGCTCACAGATGATCGATCCCTTTACGACAAAGCGCAGGC 655
Qy    315 GGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAG 374
Db    654 TGTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGGAAACTTGCCAGTGGAGCCGCG 595
Qy    375 GCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAA 434
Db    594 GCCATATTTTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGAT 535
Qy    435 GAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCA 494
Db    534 ACCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGT 475
Qy    495 AGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 551
Db    474 AGCCAGGTACCCTCTTACCATCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGA 415
Qy    552 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAG------CATGCCAG 605
Db    414 GAAAAACAAACTTGCCTGTTGGATCGTTAGCAATGCTACATCCACTGGGACAGGTACAAG 355
Qy    606 GGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATT 665
Db    354 AGCGCAATTTTATAATGAACTATCCAAACATATTAACATTAGTGTTTTTGGTGTGGTGTA 295
Qy    666 CGGAGAATACGTGAACGATAAAAATCTGATT---CCCACCATATCTACTTGTAAATTTTA 722
Db    294 CACAGGGGTCAGGTTGGAGATTGATCAGTATTACCCGACCGTTGCTAGCTGTAAGTTTTA 235
Qy    723 TCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATT 782
Db    234 CTTATCATTTGAGAACTCATTACACAAAGATTACATCACAGAGAAGGTCAACGNACCCCT 175
Qy    783 TTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATAT 842
Db    174 TGCAGCAGGAACCGTTCCTGTAGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTT 115
Qy    843 TCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATA 902
Db    114 CCCTTCAGACTCCTTTATTCACGTAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCA 55
Qy    903 TCTGAAGGAAGTTGACAA-AAACAATAAGTTGTACCTTAGTTACTTTAACTGGA 955
Db     54 TCTGCTCTCCTTGGACAAGGNATGAGGAGGCATACATGCGCTACTTTGAGTGGA 1

RESULT 13
CNS01V8V
LOCUS       CNS01V8V                  844 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            198K09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL168808
VERSION     AL168808.1  GI:7806865
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="198K09"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG198AF05SP1~end
                     PUC-Ori"
```

```
BASE COUNT      199 a    190 c    205 g    245 t      5 others
ORIGIN

  Query Match           16.3%;  Score 175.8;  DB 12;  Length 844;
  Best Local Similarity   54.4%;  Pred. No. 2.3e-38;
  Matches  398;  Conservative    2;  Mismatches  320;  Indels   12;  Gaps    2;

Qy      26 TTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTT 85
           | || || ||| | || |      | |  ||||  ||      | |   |  | ||
Db     125 TGCGGCCCTTTTTCATTGCTGTTTTTGTTTTGGGGGCCTGTGGAACTTTAATCTTTATGC 184

Qy      86 ATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGA 145
             |   || ||  |||   ||||| | ||  |||| |||| |  ||     |     | |
Db     185 TTTGGAAACCATCCACTCGCTGGCTGTTAGGTCCTGTGGAAACAACAGCAGCACCTGTYA 244

Qy     146 AAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTT 205
            |   |  | || | ||   ||          ||   |||   || ||  | ||| | |
Db     245 GAGAAATCATTTCCCTCAGGAC---------GGATAAAAACCTGACCACGGTGCTGCTCT 295

Qy     206 GGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCC 265
           || | ||||||||||||   |||||||||| |     |||  ||    |    |||
Db     296 GGTTCTGGCCATTTGGGATTACCTTTGACCCGAACGTGTGCAGCGCTTACTATGGTATAG 355

Qy     266 AAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTC 325
           | || ||     | ||| | ||| |    |  | || | ||        | || | ||||
Db     356 ACGGCTGTTTCATTACAGCGGACAGGAACTATTTCAGCGAAGGAATGCGGNTCGTTATTC 415

Qy     326 ACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTC 385
           | ||| ||||:||       || ||| | ||  | || |||||||   ||||   |  |
Db     416 ATCATCGAGAMATTTCTCCCGACCTGTCCAATCTGCCGCAGCAGGAGCGGCCGTTGTACC 475

Qy     386 AGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTG 445
           |:|  ||| | ||| | ||  | |||   ||  |  || ||  |        ||| || |
Db     476 ARATGTGGGTCTGGTTCAACATGGAGAATCCTTCGTACTCCGACAGGTTTCTTGGAATGG 535

Qy     446 AACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATG 505
           |  ||||||||||||||||| | |||||| |  ||| | |||||     ||||| |||
Db     536 ACGGCTTGTTCAACCTGACTCTGAATTATCGTCTGGATGCTGATATTTCCATGCCTCATG 595

Qy     506 GCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGG 565
           | | ||         | || |      |  ||||     || ||||| || |||||||
Db     596 TGTACCTGGTTCCTGAAAAAAGTGAGGAAGACTTTGTCCCCCCCCAGCAAAGATAAGTTGG 655

Qy     566 TGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGC 625
           | |||||| | || || ||||||||      |  ||    |||| || || | ||| ||||
Db     656 TCTGCTGGATAGTCAGCAACTGGAAGGAGCATTATATACGGGTGAAATACTTCAAGGAGC 715

Qy     626 TCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATA 685
           |  ||||   ||  | |||  ||| ||||| |  |  || || |  |||   |||
Db     716 TGTCCAAGCACATAAAGATCACCACTTATGGACGTGGCTTTGGGGGCTAC---AACTGGG 772

Qy     686 AAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTC 745
           || |       ||| |||  | |  ||||| || || || || ||||| ||||| ||
Db     773 GAAGTGCCTCGTCCATCATGGCCAGCTGTAAGTTCTACCTGTCTTTTGAGAACTCCATCT 832

Qy     746 ACAAAGATTACA 757
           |||  || ||||
Db     833 ACATTGACTACA 844


RESULT 14
BM082484/c
LOCUS       BM082484                 570 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223
            3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;,
            mRNA sequence.
ACCESSION   BM082484
VERSION     BM082484.1  GI:16929414
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN

  Query Match           15.9%;  Score 171.4;  DB 10;  Length 570;
  Best Local Similarity   57.8%;  Pred. No. 3.5e-37;
  Matches  329;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    1;

Qy     382 TTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGC 441
           || ||||| |||||||||| |  ||||||| || || ||  ||||      ||    |||
Db     570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy     442 ATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCT 501
           | ||| || |||||||||  ||||| |  ||  | | .||| | ||||| | ||| |
Db     510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy     502 TATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAG 561
             | ||    ||  || | |    || || ||||   | || |  ||||| ||
Db     450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy     562 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGGGTCAAG 612
           | || || ||| ||||||||||      |||          ||     || |||  || |
Db     390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331

Qy     613 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 672
             |||    ||  | |||       ||  |  | |||   |  ||  | || |    |  |
Db     330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271

Qy     673 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 732
           ||  |  || |||| |       ||| ||| |  |||  || |||||||| || || ||
Db     270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211

Qy     733 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 792
           || ||||| || ||||| || |||||||| || ||| || |   | |  ||   || ||
Db     210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151
```

```
Qy     793 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 852
Db     150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91

Qy     853 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 912
Db      90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31

Qy     913 GTTGACAAAAACAATAAGTTGTACCTTAG 941
Db      30 CTAAATAAAAATGACATTGCATATCACAG 2
```

```
RESULT 15
CNS02Q0S/c
LOCUS       CNS02Q0S                 887 bp    DNA     linear   GSS 14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            156L20 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL208693
VERSION     AL208693.1  GI:7867512
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 887)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .887
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="156L20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG156DF10LP1-end : T7"
BASE COUNT       232 a    188 c    206 g    253 t      8 others
ORIGIN
```

```
  Query Match             15.3%;  Score 164.8;  DB 12;  Length 887;
  Best Local Similarity   54.8%;  Pred. No. 2.9e-35;
  Matches  385;  Conservative    5;  Mismatches  297;  Indels   15;  Gaps    3;

Qy     195 CATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAAT 254
Db     704 CATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTGGCATACGACTCTTGTAAAGTTGA 645

Qy     255 GTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGC 314
Db     644 CKACAATATTGACAGCTGTGAGCTSACAGATGATCGATCCCTTGACGACAAAGCGCAGGC 585

Qy     315 GGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAG 374

Db     584 TGTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGCAAACTTGCCAGGGGAGCCGCG 525

Qy     375 GCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAA 434
Db     524 GCCATATTKTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGAT 465

Qy     435 GAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCA 494
Db     464 ACCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGT 405

Qy     495 AGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 551
Db     404 GGCCAGGTACCCTCTTACCATCAGAGAGGAAAGACTGACGGAGAAAATAGKTCTGCCGGA 345

Qy     552 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCA- 610
Db     344 GAAAAACAAAATTGTGTGCTGGATCGTCAGCAAAACTGCAAAAAASACTGGCACGGGCAC 285

Qy     611 --------AGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGC 662
Db     284 AAGGGAACAATTCTTTGGTGAACTATCCAAGCATATTCATGTTGATGTTTTTGGTGTGGC 225

Qy     663 ATTCGGAGAATACGTGAACGATAAAAATCTGATT---CCCACCATATCTACTTGTAAATT 719
Db     224 GGACACAGGAGTGAAGTTGGAGATTGATCAGTATTACCCGACCATTGCTAGCTGTAAGTT 165

Qy     720 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 779
Db     164 TTACCTGTCCTTTGAGAACTCATAACACAAAGATKACATCACAGAGAAGGTCAACGGACC 105

Qy     780 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 839
Db     104 CCTTGCAGTAGGAACCGTTCCTGTGGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTT 45

Qy     840 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTC 881
Db      44 CTTCCCTTCAGACTCCTTAATTCACGTAGATGATTTTCTTTC 3
```

```
Search completed: October  6, 2002, 07:15:31
Job time : 912.255 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 22.0361 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-3_COPY_115_1194
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128.6	11.9	1814	2	US-08-483-151-1	Sequence 1, Appli
2	128.6	11.9	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	128.2	11.9	1086	1	US-07-914-281-12	Sequence 12, Appl
4	128.2	11.9	1086	1	US-08-393-246-12	Sequence 12, Appl
5	128.2	11.9	1086	1	US-08-525-058A-12	Sequence 12, Appl
6	128.2	11.9	1086	2	US-08-696-731-12	Sequence 12, Appl
7	128.2	11.9	1086	4	US-09-042-531-12	Sequence 12, Appl
8	128.2	11.9	2042	3	US-09-063-237-2	Sequence 2, Appli
9	128.2	11.9	2043	1	US-07-914-281-1	Sequence 1, Appli
10	128.2	11.9	2043	1	US-08-393-246-1	Sequence 1, Appli
11	128.2	11.9	2043	1	US-08-273-411-4	Sequence 4, Appli
12	128.2	11.9	2043	1	US-08-525-058A-1	Sequence 1, Appli
13	128.2	11.9	2043	2	US-08-696-731-1	Sequence 1, Appli
14	128.2	11.9	2043	4	US-09-042-531-1	Sequence 1, Appli
15	128.2	11.9	2043	5	PCT-US91-00899-1	Sequence 1, Appli
16	126.6	11.7	1316	1	US-07-914-281-10	Sequence 10, Appl
17	126.6	11.7	1316	1	US-08-393-246-10	Sequence 10, Appl
18	126.6	11.7	1316	1	US-08-525-058A-10	Sequence 10, Appl
19	126.6	11.7	1316	2	US-08-696-731-10	Sequence 10, Appl
20	126.6	11.7	1316	4	US-09-042-531-10	Sequence 10, Appl
21	124.6	11.5	1654	1	US-07-914-281-13	Sequence 13, Appl
22	124.6	11.5	1654	1	US-08-393-246-13	Sequence 13, Appl
23	124.6	11.5	1654	1	US-08-525-058A-13	Sequence 13, Appl
24	124.6	11.5	1654	2	US-08-696-731-13	Sequence 13, Appl
25	124.6	11.5	1654	4	US-09-042-531-13	Sequence 13, Appl
26	120	11.1	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	120	11.1	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	120	11.1	1488	1	US-07-914-281-9	Sequence 9, Appli
29	120	11.1	1488	1	US-08-393-246-9	Sequence 9, Appli
30	120	11.1	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	120	11.1	1488	2	US-08-696-731-9	Sequence 9, Appli
32	120	11.1	1488	4	US-09-042-531-9	Sequence 9, Appli
33	120	11.1	2175	4	US-08-482-073-9	Sequence 9, Appli
34	120	11.1	2861	4	US-08-482-073-10	Sequence 10, Appl
35	120	11.1	3647	1	US-07-914-281-7	Sequence 7, Appli
36	120	11.1	3647	1	US-08-393-246-7	Sequence 7, Appli
37	120	11.1	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	120	11.1	3647	2	US-08-696-731-7	Sequence 7, Appli
39	120	11.1	3647	4	US-09-042-531-7	Sequence 7, Appli
40	120	11.1	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	116.8	10.8	2134	2	US-08-483-151-3	Sequence 3, Appli
42	116.8	10.8	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	33.8	3.1	1730	1	US-07-817-920-1	Sequence 1, Appli
c 44	33.8	3.1	1730	1	US-08-117-006-1	Sequence 1, Appli
c 45	33.8	3.1	1730	1	US-08-216-594-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
;  GENERAL INFORMATION:
;    APPLICANT:  Seed, Brian
;    APPLICANT:  Holgersson, Jan
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-151-1

  Query Match           11.9%;  Score 128.6;  DB 2;  Length 1814;
  Best Local Similarity  52.5%;  Pred. No. 1.1e-29;
  Matches  331;  Conservative    0;  Mismatches  294;  Indels    6;  Gaps    2;

Qy    269 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 328
          | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy     329 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 388
           |  | ||  |          |   ||  | |||| | | |   |||||||  |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     389 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 448
           |||  | |||       | ||  || |||| | | ||||          | || |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     449 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           | | ||||||  |   || |  ||||| |||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     509 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 568
           ||||  |     | |   |||       |    | | ||  |||    | | |||
Db     788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 628
           ||||||  | || || |   |       |||      | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     629 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 688
           |     | |  | |   |    |  || |  ||   ||||     | |  ||| |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     689 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 748
           ||||  | |||||  |  |     |      || || ||  | ||||| ||||||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     749 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 805
           ||  |||||||| || ||| |||     ||||||  |   ||||||| | ||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     806 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 865
           |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     866 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 896
           |||| || || | |||| || ||||  | ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

```
RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;     APPLICANT:  The General Hospital Corporation
;     TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/06427
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/483,151
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lech, Karen F.
;       REGISTRATION NUMBER:  35,238
;       REFERENCE/DOCKET NUMBER:  00786/278WO1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1814 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             11.9%;  Score 128.6;  DB 5;  Length 1814;
  Best Local Similarity   52.5%;  Pred. No. 1.1e-29;
  Matches  331;  Conservative    0;  Mismatches  294;  Indels    6;  Gaps    2;

Qy     269 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 328
           | |||| ||| |    |  ||||     ||      ||   |  |||| ||  |  | ||||
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     329 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 388
           |  | ||  |           |   ||  | |||| | | |    |||||||  |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     389 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 448
           |||  | |||       | ||  || |||| | | ||||          | || |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     449 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           | | ||||||  |   || |  ||||| |||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     509 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 568
           ||||  |     | |   |||       |    | | ||  |||    | | |||
Db     788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 628
           ||||||  | || || |   |       |||      | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     629 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 688
           |     | |  | |   |    |  || |  ||   ||||     | |  ||| |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     689 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 748
           ||||  | |||||  |  |     |      || || ||  | ||||| ||||||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     749 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 805
           ||  |||||||| || ||| |||     ||||||  |   ||||||| | ||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     806 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 865
           |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     866 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 896
           |||| || || | |||| || ||||  | ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

```
RESULT 3
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;     APPLICANT:  LOWE, JOHN B.
;     TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;     TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;     TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;     TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match            11.9%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 1.2e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
          ||||| |||| | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
          | || ||||  |   |  | | |  | | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
          |||||  ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
          |   ||  | | |  |    |   ||          | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
          | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
          ||  | |||   || ||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
          || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
          |||||  || |||||  || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 4
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match            11.9%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 1.2e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
          ||||| |||| | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
          | || ||||  |   |  | | |  | | || |      | |||||  |     |||
```

```
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447

Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507

Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 5
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;    INFORMATION FOR SEQ ID NO:  12:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  1086 base pairs
;        TYPE:  nucleic acid
;        STRANDEDNESS:  double
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match
  Best Local Similarity   11.9%;  Score 128.2;  DB 1;  Length 1086;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330

Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387

Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447

Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507

Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 6
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
```

```
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             11.9%;  Score 128.2;  DB 2;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 1.2e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
          | || ||||  |   |   | | |   | | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
            |||||.||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
          |   ||  |  | |  |   |   ||           | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
            || |||||  |||||  |||   |||||||| || ||    ||||||||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
            | |    |  ||||   ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
          || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
          ||||| || ||||| || ||  |    ||||  ||  | ||     || ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 7
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match                 11.9%;  Score 128.2;  DB 4;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 1.2e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330
Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390
Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450
Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510
Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570
Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630
Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687
Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747
Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807
Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867
Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927
Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 8
US-09-063-237-2
, Sequence 2, Application US/09063237
, Patent No. 6124267
, GENERAL INFORMATION:
,   APPLICANT:  McEver, Rodger P.
,   APPLICANT:  Cummings, Richard D.
,   TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
,   TITLE OF INVENTION:  Inflammation Derived from PSGL-1
,   NUMBER OF SEQUENCES:  5
,   CORRESPONDENCE ADDRESS:
,     ADDRESSEE:  Patrea L. Pabst
,     STREET:  2800 One Atlantic Center, 1201 West Peachtree
,     STREET:  Street
,     CITY:  Atlanta
,     STATE:  Georgia
,     COUNTRY:  US
,     ZIP:  30306-3450
,   COMPUTER READABLE FORM:
,     MEDIUM TYPE:  Floppy disk
,     COMPUTER:  IBM PC compatible
,     OPERATING SYSTEM:  PC-DOS/MS-DOS
,     SOFTWARE:  PatentIn Release #1.0, Version #1.25
,   CURRENT APPLICATION DATA:
,     APPLICATION NUMBER:  US/09/063,237
,     FILING DATE:
,     CLASSIFICATION:
,   PRIOR APPLICATION DATA:
,     APPLICATION NUMBER:  08/649,802
,     FILING DATE:
,   ATTORNEY/AGENT INFORMATION:
,     NAME:  Pabst, Patrea L.
,     REGISTRATION NUMBER:  31,284
,     REFERENCE/DOCKET NUMBER:  OMRF110CIP7
,   TELECOMMUNICATION INFORMATION:
,     TELEPHONE:  (404)873-8794
,     TELEFAX:  (404)873-8795
, INFORMATION FOR SEQ ID NO:  2:
,   SEQUENCE CHARACTERISTICS:
,     LENGTH:  2042 base pairs
,     TYPE:  nucleic acid
,     STRANDEDNESS:  double
,     TOPOLOGY:  linear
,   MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match                 11.9%;  Score 128.2;  DB 3;  Length 2042;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402
Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462
Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522
Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582
Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642
Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702
Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759
Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819
Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879
Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
```

```
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
          ||||| || ||||| || ||  |    |||| ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||  || |      ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 9
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             11.9%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
          | || ||||  |   |   |  | |   | | || |      | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
          |||||  ||| | || || |||  ||||| || ||  ||  ||   |  | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
          |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
          || |||||  |||||  |||  |||||||| || ||    ||||||||     || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
          |    ||         | |  |  |  ||   || || |   |   |   |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
          | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
          || || || || ||||| ||    ||    ||||| |||| |   |  | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
          ||||| || ||||| || ||  |    |||| ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
          |||||  || |      ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 10
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
```

```
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             11.9%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           ||||| ||||  | ||||||     |||||     |  |  |  ||||| |   | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
            || |||| |   |   |  | | |   | | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
              ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
            || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |   ||       | |  |  |  ||   || || |   |   |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
           | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
           ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
           || || || || ||||| ||    ||    |||||  |||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
           ||||| || ||||| || ||   |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 11
US-08-273-411-4
; Sequence 4, Application US/08273411
; Patent No. 5625124
;  GENERAL INFORMATION:
;    APPLICANT:  Falk, Per
;    APPLICANT:  Gordon, Jeffrey I.
;    TITLE OF INVENTION:  Animal Model for Gastro-Intestinal
;    TITLE OF INVENTION:  Disease
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/273,411
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 815-6508
;      TELEFAX:  (404) 815-6555
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..361
;      OTHER INFORMATION:  /note= "Nucleotides 73 through 1158 encode the GDP-L-fuco
;    PUBLICATION INFORMATION:
;      AUTHORS:  Kukowska-Latallo, et al.
;      JOURNAL:  Genes & Development
;      VOLUME:  4
;      PAGES:  1288-1303
;      DATE:  1990
;      RELEVANT RESIDUES IN SEQ ID NO:  4: FROM 1 TO 2043
US-08-273-411-4

  Query Match             11.9%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           ||||| ||||  | ||||||     |||||     |  |  |  ||||| |   | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
            || |||| |   |   |  | | |   | | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
              ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642
```

```
Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
           || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |    ||          | |  |  |  ||  || || |   |   |     |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
           | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
           ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
           || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 12
US-08-525-058A-1
; Sequence 1, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,058A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO

US-08-525-058A-1

Query Match 11.9%; Score 128.2; DB 1; Length 2043;
Best Local Similarity 53.0%; Pred. No. 1.6e-29;
Matches 374; Conservative 0; Mismatches 313; Indels 18; Gaps 4;

```
Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           |||||  ||||  | ||||||    |||||    |  |  |  |||||  |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |  |   |  | |   | | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
           |||||  |||  | || || |||  ||||| || ||  ||  ||  |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
           |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |     |   ||          | |  |  |  |  ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
           || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |    ||          | |  |  |  ||  || || |   |   |     |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
           | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
           ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
           || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 13
US-08-696-731-1
; Sequence 1, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.

```
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/696,731
;       FILING DATE:  14-AUG-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2043 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     ANTI-SENSE:  NO
US-08-696-731-1

  Query Match             11.9%;  Score 128.2;  DB 2;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
           ||||| |||| | |||||    |||||     |  |  |  ||||| |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387
           | || ||||  |   |   |  | |   | | || |      | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447
           |||||  ||| | || || ||| |||||  || ||   ||   ||    |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558
           |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618
           || |||||  |||||  |||   |||||||| || ||    |||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678
           |    ||         | |  |  |  ||   || || |   |   |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738
           | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795
           ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
```

```
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855
           || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-09-042-531-1
; Sequence 1, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-09-042-531-1

  Query Match             11.9%;  Score 128.2;  DB 4;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330
```

```
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 15
PCT-US91-00899-1
; Sequence 1, Application PC/TUS9100899
;  GENERAL INFORMATION:
;    APPLICANT:  Lowe, John B.
;    TITLE OF INVENTION:  Method and Products For the Synthesis of
;    TITLE OF INVENTION:  Oligosaccharide Structures on Glycoproteins, Glycolipids,
;    TITLE OF INVENTION:  or as Free Molecules, and For the Isolation of Cloned
;    TITLE OF INVENTION:  Genetic Sequences That Determine These Structur
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00899
;      FILING DATE:  19910214
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye Ph.D., Jean-Paul
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-021-55 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-5940
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Blood
;      CELL LINE:  A431
PCT-US91-00899-1

  Query Match             11.9%;  Score 128.2;  DB 5;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-29;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 330

Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 387

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 447

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 558

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 618

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 678

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    679 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 738

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 795

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 855

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 915

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 960
```

```
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:20:34
Job time : 29.0361 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 1146.59 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-3_COPY_280_1194
Perfect score: 915
Sequence: 1 acaaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	915	100.0	2139	10	AB015426	AB015426 Mus muscu
2	856	93.6	1128	10	AF345993	AF345993 Rattus no
3	852.8	93.2	2156	10	AB049819	AB049819 Rattus no
4	825.4	90.2	1705	10	AF230460	AF230460 Cricetulu
5	796.6	87.1	2501	9	HSA238701	AJ238701 Homo sapi
6	793.4	86.7	80247	9	AL512406	AL512406 Human DNA
7	791.8	86.5	3019	9	AB023021	AB023021 Homo sapi
8	668.4	73.0	71396	2	AC100674	AC100674 Mus muscu
9	629.2	68.8	1707	5	AB035906	AB035906 Gallus ga
10	589.2	64.4	1080	5	AB035905	AB035905 Xenopus l
11	268.2	29.3	1567	5	AB023627	AB023627 Danio rer
12	228.2	24.9	1832	5	AB023628	AB023628 Danio rer
13	220.6	24.1	1264	5	AF288369	AF288369 Gallus ga
14	206.6	22.6	1092	5	AB035907	AB035907 Xenopus l
15	175.2	19.1	2657	5	GGU73678	U73678 Gallus gall
16	157	17.2	1429	10	CGU78737	U78737 Cricetulus
17	157	17.2	1462	10	AF090450	AF090450 Cricetulu
18	155.4	17.0	1718	10	AF090449	AF090449 Cricetulu
c 19	145.6	15.9	71396	2	AC100674	AC100674 Mus muscu
20	135.6	14.8	1044	5	AB035908	AB035908 Gallus ga
21	134.8	14.7	2293	4	BTA134FTG	X87810 B.taurus al
22	134.8	14.7	12607	4	BTA132772	AJ132772 Bos tauru
23	134.2	14.7	1158	10	AB039126	AB039126 Mus muscu
24	133.4	14.6	1113	9	AF345884	AF345884 Saimiri s
25	133	14.5	1119	9	AF345881	AF345881 Macaca mu
26	132.6	14.5	1170	10	AB039127	AB039127 Mus muscu
27	132.6	14.5	1170	10	AB039131	AB039131 Mus muscu
28	132.6	14.5	1615	10	MMU33457	U33457 Mus musculu
29	131.4	14.4	1170	10	AB039133	AB039133 Mus spici
30	131	14.3	1170	10	AB039124	AB039124 Mus muscu
31	131	14.3	1170	10	AB039125	AB039125 Mus muscu
32	131	14.3	1170	10	AB039128	AB039128 Mus muscu
33	131	14.3	1170	10	AB039129	AB039129 Mus muscu
34	131	14.3	1170	10	AB039130	AB039130 Mus muscu
35	131	14.3	1170	10	AB039132	AB039132 Mus muscu
36	130.8	14.3	1302	10	AB049938	AB049938 Rattus no
37	130.8	14.3	169446	2	AC097778	AC097778 Rattus no
38	130	14.2	1613	4	BTA132776	AJ132776 Bos tauru
39	130	14.2	1619	4	BTA132775	AJ132775 Bos tauru
40	130	14.2	1644	4	BTA132774	AJ132774 Bos tauru
41	130	14.2	1873	4	BTA132773	AJ132773 Bos tauru
42	129.8	14.2	1162	9	PTY14033	Y14033 Pan troglod
c 43	129.8	14.2	80893	9	AC024592	AC024592 Homo sapi
44	129.8	14.2	187488	2	AC010454	AC010454 Homo sapi
c 45	128.8	14.1	169893	2	AC021165	AC021165 Homo sapi

ALIGNMENTS

RESULT 1
AB015426
LOCUS AB015426 2139 bp mRNA linear ROD 23-JUN-1999
DEFINITION Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete cds.
ACCESSION AB015426
VERSION AB015426.1 GI:3702718
KEYWORDS Fut9; alpha1,3-fucosyltransferase IX.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K. and Narimatsu,H.
TITLE Expression cloning and characterization of a novel murine alpha1, 3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15) epitope in brain and kidney
JOURNAL J. Biol. Chem. 273 (41), 26729-26738 (1998)
MEDLINE 98434588
REFERENCE 2 (bases 1 to 2139)
AUTHORS Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN

  Query Match             100.0%;  Score 915;  DB 10;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-233;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 339

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     340 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 399

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     400 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 459

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 519

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 579

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 639

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 699

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 759

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 819

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 879

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 939

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 999

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 1059

Qy     781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1119

Qy     841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1179

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    1180 TGGTTTTGGAATTAA 1194


RESULT 2
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN

  Query Match             93.6%;  Score 856;  DB 10;  Length 1128;
```

```
  Best Local Similarity   96.2%;  Pred. No. 4.8e-217;
  Matches  877;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       4 AAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACC 63
           |||||||||||||||| |||||||| |||||||||||||| ||||||||||||||||||
Db     217 AAAACTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACC 276

Qy      64 TTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGAC 123
           ||||||||||||||||||||||| |||||||| ||||||||||||||||||||||| |||
Db     277 TTTGACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGAC 336

Qy     124 CGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGAT 183
           ||||| |||||||||||||||||||||||||||| ||||| ||||||||||||||||||
Db     337 CGCTCGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGAT 396

Qy     184 CTGACTAACTTACCTCAGCAGGCCAGGCCAGCCCTTTCAGAAATGGATTTGGATGAATTTA 243
           |||||||||||||||||||||||||||| |||||||||||||||||||||||| |||
Db     397 CTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTA 456

Qy     244 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTA 303
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     457 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTA 516

Qy     304 ACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAAT 363
           |||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||
Db     517 ACTTATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGC 576

Qy     364 CCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGG 423
           |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db     577 CCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGG 636

Qy     424 AACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 483
           || || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     637 AATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 696

Qy     484 ACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCT 543
           |||||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||
Db     697 ACCTATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCT 756

Qy     544 ACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 603
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     757 ACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 816

Qy     604 CTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAAC 663
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     817 CTCTACAATGCATTTTTGGCTGGTTCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAAC 876

Qy     664 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGT 723
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     877 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGT 936

Qy     724 GAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTT 783
           ||||| |||||||||||||||| |||||||||||||||||| || ||||||||||||
Db     937 GAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTC 996

Qy     784 AACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTG 843
           |||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     997 AACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTG 1056

Qy     844 GCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGG 903
           |||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    1057 GCATGCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGG 1116

Qy     904 TTTTGGAATTAA 915
           ||||||||||||
Db    1117 TTTTGGAATTAA 1128


RESULT 3
AB049819
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN

  Query Match             93.2%;  Score 852.8;  DB 10;  Length 2156;
  Best Local Similarity   95.9%;  Pred. No. 3.4e-216;
  Matches  875;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       4 AAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACC 63
           |||||||||||||||| |||||||| |||||||||||||| ||||||||||||||||||
Db     514 AAAACTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACC 573

Qy      64 TTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGAC 123
           ||||||||||||||||||||||| |||||||| ||||||||||||||||||||||| |||
Db     574 TTTGACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGAC 633

Qy     124 CGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGAT 183
           |||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||
Db     634 CGCTCGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGAT 693

Qy     184 CTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTA 243
           |||||||||||||||||||||||||||| |||||||||||||||||||||||| |||
Db     694 CTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTA 753

Qy     244 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTA 303
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     754 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTA 813

Qy     304 ACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAAT 363
           |||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||
Db     814 ACTTATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGC 873
```

```
Qy     364 CCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGG 423
           ||||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db     874 CCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGG 933

Qy     424 AACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 483
           || || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     934 AATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 993

Qy     484 ACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCT 543
           |||||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||
Db     994 ACCTATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCT 1053

Qy     544 ACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 603
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1054 ACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 1113

Qy     604 CTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAAC 663
           |||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||
Db    1114 CTCTACAATGCATTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAAC 1173

Qy     664 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGT 723
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1174 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGT 1233

Qy     724 GAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTT 783
           ||||| |||||||||||||||| |||||||||||||||||||| || ||||||||||||
Db    1234 GAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTC 1293

Qy     784 AACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTG 843
           |||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    1294 AACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTG 1353

Qy     844 GCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGG 903
           |||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1354 GCATGCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGG 1413

Qy     904 TTTTGGAATTAA 915
           ||||||||||||
Db    1414 TTTTGGAATTAA 1425
```

```
RESULT 4
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPPQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN
```

```
  Query Match             90.2%;  Score 825.4;  DB 10;  Length 1705;
  Best Local Similarity   93.9%;  Pred. No. 6.7e-209;
  Matches  859;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
           |||||||||||||||||||| ||||| || ||||| |||||||| ||||||||||||||||
Db     227 ACAAAAACTGATTATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAG 286

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
           |||||||||||||||||||||||||| ||||||| |||||||||||||||||||| |||
Db     287 ACCTTTGACCTTACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACA 346

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
           ||||||||| ||||||||||||||||||||||||||||||||| ||||| ||||||||||||
Db     347 GACCGCTCACTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGG 406

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
           |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db     407 GATCTGACTAACTTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAAC 466

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
           ||||||| |||||||| |||||||||||||||||||||||| || ||||||||||||||
Db     467 CTAGAGTCTCCCACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACT 526

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 CTAACATATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 586

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||
Db     587 AATCCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAAC 646

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
           ||||| || |||||||||||||||||||||||| || ||||||||||||||||||||||||
Db     647 TGGAATCCGGAGCATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATC 706

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
           || |||||||| ||||| || |||||||||||||| || |||||||| ||||||||||||
Db     707 CATACCTATGGGCAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATA 766

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||
Db     767 TCTACTTGTAAATTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAA 826

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
           |||||||||||||||||||||||||| |||||||||||||| ||||| ||||| ||||||
Db     827 AAGCTCTACAATGCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAA 886

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
           |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Db     887 AACTATGAGAATTATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 946

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
           |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
```

```
Db    947 AGTGAGTTGGCAAAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAT 1006

Qy    781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
          || |||||||| |||||||| |||||||||| ||||| ||||||||||||||||||||
Db   1007 TTCAACTGGAGGAAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGC 1066

Qy    841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||| || || || |||||||||| ||||| |||||||||| ||||||||||||||||
Db   1067 CTGGCTTGTGATCACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAA 1126

Qy    901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db   1127 TGGTTTTGGAATTAA 1141


RESULT 5
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental

BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match            87.1%;  Score 796.6;  DB 9;  Length 2501;
  Best Local Similarity  91.9%;  Pred. No. 3.1e-201;
  Matches  841;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy      1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          || |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db    270 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 329

Qy     61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db    330 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 389

Qy    121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
          ||||| ||| |||||||||||||||||| || ||||| || ||  ||||||||||||||
Db    390 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGCTGG 449

Qy    181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
          |||||||||||||||||||||| || |||||||||| ||||||||||||||||||||||
Db    450 GATCTGACTAACTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 509

Qy    241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
          || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db    510 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 569

Qy    301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
          || ||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    570 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 629

Qy    361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
          |||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||
Db    630 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 689

Qy    421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db    690 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 749

Qy    481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
          || ||||| || |||||||| |||||||| || || |||||||| ||||||| ||||||
Db    750 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 809

Qy    541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
          ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db    810 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 869

Qy    601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
          ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db    870 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 929

Qy    661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
          |||||||||||||||||||||||| ||||||||||||||||||||||||| ||||||||||
Db    930 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 989

Qy    721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
          |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db    990 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1049

Qy    781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
          |||||||||| |||||||| ||||||| || ||||| ||||||||||||||||||||
Db   1050 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1109

Qy    841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1110 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1169

Qy    901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db   1170 TGGTTTTGGAATTAA 1184


RESULT 6
```

```
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-504J9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-381A2 is at 80148 in this sequence.
            The true right end of clone RP11-77B15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .80247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-504J9"
                     /clone_lib="RPCI-11.2"
     repeat_region   184. .1422
                     /note="L2 repeat: matches 875. .2419 of consensus"
     repeat_region   1744. .1834
                     /note="MIR repeat: matches 166. .258 of consensus"
     repeat_region   2950. .2977
                     /note="14 copies 2 mer ca 100% conserved"
     repeat_region   3411. .3669
                     /note="LTR16C repeat: matches 93. .364 of consensus"
     repeat_region   4372. .4704
                     /note="MLT1A1 repeat: matches 5. .365 of consensus"
     misc_feature    complement(7821. .8406)
                     /note="match: GSS: Em:AQ550748"
     repeat_region   8126. .8184
                     /note="MER58B repeat: matches 272. .332 of consensus"
     repeat_region   8185. .8696
                     /note="MLT2D repeat: matches 1. .513 of consensus"
     repeat_region   8779. .8817
                     /note="MLT2D repeat: matches 504. .542 of consensus"
     repeat_region   8819. .8974
                     /note="MER58A repeat: matches 6. .162 of consensus"
     misc_feature    complement(9443. .9891)
                     /note="match: GSS: Em:B90633"
     misc_feature    complement(9452. .9883)
                     /note="match: GSS: Em:AQ119695"
     repeat_region   9782. .9894
                     /note="MIR repeat: matches 92. .205 of consensus"
     repeat_region   10089. .10212
                     /note="MIR repeat: matches 135. .262 of consensus"
     misc_feature    complement(10093. .10671)
                     /note="match: GSS: Em:AQ237061"
     repeat_region   10223. .10282
                     /note="MIR repeat: matches 203. .262 of consensus"
     misc_feature    complement(10315. .10631)
                     /note="match: GSS: Em:AQ200259"
     misc_feature    10805. .11186
                     /note="match: GSS: Em:AQ072353"
     repeat_region   11657. .11844
                     /note="L2 repeat: matches 2546. .2750 of consensus"
     misc_feature    complement(11659. .12084)
                     /note="match: GSS: Em:AQ518361"
     misc_feature    complement(11674. .11987)
                     /note="match: GSS: Em:AQ094678"
     misc_feature    12083. .12588
                     /note="match: GSS: Em:AQ691235"
     repeat_region   12596. .12810
                     /note="MIR repeat: matches 23. .248 of consensus"
     repeat_region   12920. .12986
                     /note="L1MA5 repeat: matches 6228. .6295 of consensus"
     repeat_region   12987. .13169
                     /note="L1MA5 repeat: matches 5852. .6043 of consensus"
     repeat_region   13168. .13280
                     /note="L1MA8 repeat: matches 5710. .5832 of consensus"
     repeat_region   13851. .17218
                     /note="L1PB1 repeat: matches 2787. .6155 of consensus"
     repeat_region   18903. .18953
                     /note="MLT2CB repeat: matches 439. .501 of consensus"
     repeat_region   18984. .19433
                     /note="MLT2CB repeat: matches 1. .460 of consensus"
     misc_feature    complement(19108. .19690)
                     /note="match: GSS: Em:AQ563935"
     repeat_region   20683. .20801
                     /note="AluJo/FRAM repeat: matches 169. .288 of consensus"
     repeat_region   20802. .20862
                     /note="AluJ/FLAM repeat: matches 3. .66 of consensus"
     repeat_region   20912. .21033
                     /note="MIR repeat: matches 114. .237 of consensus"
     repeat_region   21047. .21128
                     /note="41 copies 2 mer ta 63% conserved"
     repeat_region   21676. .21859
                     /note="L2 repeat: matches 2130. .2308 of consensus"
     repeat_region   21890. .23612
                     /note="L1P repeat: matches 841. .2570 of consensus"
     repeat_region   23601. .26609
                     /note="L1PA5 repeat: matches 3107. .6145 of consensus"
     repeat_region   26620. .26789
                     /note="MER5A repeat: matches 1. .174 of consensus"
     repeat_region   28521. .28627
                     /note="MER39 repeat: matches 5. .111 of consensus"
     repeat_region   28656. .28837
                     /note="MER39b repeat: matches 54. .235 of consensus"
     repeat_region   28838. .29142
                     /note="AluSc repeat: matches 4. .308 of consensus"
     repeat_region   29143. .29494
                     /note="MER39b repeat: matches 235. .549 of consensus"
     repeat_region   30529. .31344
                     /note="L1PA3 repeat: matches 5331. .6146 of consensus"
     repeat_region   32234. .32340
                     /note="L2 repeat: matches 2627. .2750 of consensus"
     repeat_region   32399. .32578
                     /note="MIR repeat: matches 12. .203 of consensus"
```

```
     repeat_region   32816. .32979
                     /note="MIR repeat: matches 91. .259 of consensus"
     misc_feature    complement(33006. .33450)
                     /note="match: GSS: Em:AQ261227"
     misc_feature    complement(33058. .33444)
                     /note="match: GSS: Em:AQ040453"
     repeat_region   33312. .33596
                     /note="L1M4 repeat: matches 3473. .3773 of consensus"
     repeat_region   34856. .34915
                     /note="30 copies 2 mer aa 68% conserved"
     misc_feature    complement(34987. .35547)
                     /note="match: GSS: Em:AQ321598
                     match: STS: Em:G55571"
     misc_feature    complement(36488. .36936)
                     /note="match: GSS: Em:B41073"
     misc_feature    complement(36654. .37090)
                     /note="match: GSS: Em:AQ772584"
     misc_feature    37102. .37591
                     /note="match: GSS: Em:AQ489567"
     misc_feature    complement(37581. .38221)
                     /note="match: GSS: Em:AQ418233"
     repeat_region   37720. .37772
                     /note="L2 repeat: matches 2657. .2710 of consensus"
     misc_feature    complement(37778. .38215)
                     /note="match: GSS: Em:AQ819773"
     misc_feature    38218. .38732
                     /note="match: GSS: Em:AQ599868"
     repeat_region   38861. .39164
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   39612. .39653
                     /note="21 copies 2 mer ac 90% conserved"
     repeat_region   40854. .40909
                     /note="L2 repeat: matches 2655. .2710 of consensus"
     repeat_region   42169. .42270
                     /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   42411. .42548
                     /note="MER5B repeat: matches 21. .181 of consensus"
     repeat_region   42780. .42938
                     /note="MER5B repeat: matches 17. .177 of consensus"
     repeat_region   43229. .43326
                     /note="MIR repeat: matches 40. .148 of consensus"
     repeat_region   45383. .45428
                     /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                     /note="match: STS: Em:G51995
                     match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                     /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
                     /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                     /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                     /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match           86.7%;  Score 793.4;  DB 9;  Length 80247;
  Best Local Similarity  91.7%;  Pred. No. 2.3e-200;
  Matches  839;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
           || ||||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db   35724 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 35783

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
           |||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db   35784 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 35843

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
           ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db   35844 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 35903

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db   35904 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 35963

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db   35964 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 36023

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   36024 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 36083

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||
Db   36084 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 36143

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||||||
Db   36144 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 36203

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
           || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db   36204 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 36263

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
           |||||||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db   36264 TCTACTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 36323

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
           ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db   36324 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 36383

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
           |||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db   36384 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 36443

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
           |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db   36444 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 36503

Qy     781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
           |||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db   36504 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 36563

Qy     841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db   36564 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 36623

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db   36624 TGGTTTTGGAATTAA 36638


RESULT 7
AB023021
LOCUS       AB023021               3019 bp    mRNA    linear   PRI 23-JUN-1999
DEFINITION  Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB023021
VERSION     AB023021.1  GI:5139692
KEYWORDS    alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S.,
            Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and
            Narimatsu,H.
  TITLE     Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved
            between human and mouse; molecular cloning, characterization and
            tissue distribution of human Fuc-TIX
  JOURNAL   FEBS Lett. 453, 237-242 (1999)
REFERENCE   2  (bases 1 to 3019)
```

```
  AUTHORS   Kaneko,M., Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
ORIGIN

  Query Match           86.5%;  Score 791.8;  DB 9;  Length 3019;
  Best Local Similarity   91.6%;  Pred. No. 5.9e-200;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
Db     460 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 519

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
Db     520 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 579

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
Db     580 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 639

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
Db     640 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 699

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
Db     700 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 759

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
Db     760 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 819

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
Db     820 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 879

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
Db     880 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 939

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
Db     940 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 999

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
Db    1000 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 1059

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
Db    1060 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 1119

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
Db    1120 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 1179

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
Db    1180 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1239

Qy     781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
Db    1240 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1299

Qy     841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
Db    1300 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1359

Qy     901 TGGTTTTGGAATTAA 915
Db    1360 TGGTTTTGGAATTAA 1374


RESULT 8
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
```

```
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L15932
    Center clone name: 167_K_24
--------------.
* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      788: contig of 788 bp in length
*      789 888: gap of      100 bp
*      889     1702: contig of 814 bp in length
*     1703 1802: gap of      100 bp
*     1803     2642: contig of 840 bp in length
*     2643 2742: gap of      100 bp
*     2743     3637: contig of 895 bp in length
*     3638 3737: gap of      100 bp
*     3738     4635: contig of 898 bp in length
*     4636 4735: gap of      100 bp
*     4736     5657: contig of 922 bp in length
*     5658 5757: gap of      100 bp
*     5758     6692: contig of 935 bp in length
*     6693 6792: gap of      100 bp
*     6793     7703: contig of 911 bp in length
*     7704 7803: gap of      100 bp
*     7804     8742: contig of 939 bp in length
*     8743 8842: gap of      100 bp
*     8843     9787: contig of 945 bp in length
*     9788 9887: gap of      100 bp
*     9888    10819: contig of 932 bp in length
*    10820 10919: gap of      100 bp
*    10920    11808: contig of 889 bp in length
*    11809 11908: gap of      100 bp
*    11909    12830: contig of 922 bp in length
*    12831 12930: gap of      100 bp
*    12931    13847: contig of 917 bp in length
*    13848 13947: gap of      100 bp
*    13948    14886: contig of 939 bp in length
*    14887 14986: gap of      100 bp
*    14987    15911: contig of 925 bp in length
*    15912 16011: gap of      100 bp
*    16012    16903: contig of 892 bp in length
*    16904 17003: gap of      100 bp
*    17004    17906: contig of 903 bp in length
*    17907 18006: gap of      100 bp
*    18007    18931: contig of 925 bp in length
*    18932 19031: gap of      100 bp
*    19032    19945: contig of 914 bp in length
*    19946 20045: gap of      100 bp
*    20046    20951: contig of 906 bp in length
*    20952 21051: gap of      100 bp
*    21052    21949: contig of 898 bp in length
*    21950 22049: gap of      100 bp
*    22050    22950: contig of 901 bp in length
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
```

```
          *    65455    66352: contig of 898 bp in length
          *    66353 66452: gap of      100 bp
          *    66453    67362: contig of 910 bp in length
          *    67363 67462: gap of      100 bp
          *    67463    68370: contig of 908 bp in length
          *    68371 68470: gap of      100 bp
          *    68471    69379: contig of 909 bp in length
          *    69380 69479: gap of      100 bp

  Query Match           73.0%;  Score 668.4;  DB 2;  Length 71396;
  Best Local Similarity  83.2%;  Pred. No. 4.2e-167;
  Matches  764;  Conservative    0;  Mismatches  140;  Indels   14;  Gaps    6;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65550 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 65609

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65610 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 65669

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65670 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 65729

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65730 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 65789

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65790 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 65849

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65850 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 65909

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65910 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 65969

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65970 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 66029

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66030 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 66089

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
           |||||||||||||||||||| | |||||||||| |    | |     |   | ||
Db   66090 TCTACTTGTAAATTTTATCTTTC-TCTGAAAACTCATTCACAAGATTACTCCCAAAAG-- 66146

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
             | ||     ||||||||||||||||||||||||||||||||||||||||||||||||
Db   66147 ------TCCACATGCTTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 66200

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTC--ATGTGGAAGATTTTAACTCT- 717
           ||||||||||||||||||||||||||||||||||||    |||||||||||| |||||
Db   66201 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCCATCGTGGAAGATTTTCACTCTC 66260

Qy     718 CCCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTT-GTACCTTAG 776
           ||||||||||||||||    |||||||||| ||||| |||||||||| |||||||||
Db   66261 CCCAGTGAGTTAGCAAATATCCTGAAGGAAGTCGACAANAACAATAAGTTGGTACCTTAG 66320

Qy     777 TTACTTTAACT-GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATG 835
           ||||| |||||  | ||||||||| ||| |
Db   66321 TTACTCTAACTCGNAGAAAGGATTCTACCGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 66380

Qy     836 CATGCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAG 895

Db   66381 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 66440
```

```
Qy     896 AGAAATGGTTTTGGAATT 913
                        |||
Db   66441 NNNNNNNNNNNNNCGATT 66458


RESULT 9
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match           68.8%;  Score 629.2;  DB 5;  Length 1707;
  Best Local Similarity  80.5%;  Pred. No. 1.1e-156;
  Matches  736;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy       2 CAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGA 61
           | |||||||||| ||||||| |||||||  ||| |||||||| | ||||||||||| ||||
Db     434 CCAAAACTGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGA 493

Qy      62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAG 121
           | || || || |||||||||||| | |||||||| ||||| |||||||||| || |  |
Db     494 CATTCGATCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTG 553

Qy     122 ACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGG 181
           |||||||| | || ||||| |||||||| || || || || || || ||||| | |||||
Db     554 ACCGCTCACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGG 613

Qy     182 ATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATT 241
           |||||||||| |||||||||| |||||||||| || |||||||||||||||||||| |
Db     614 ATCTGACTAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACT 673

Qy     242 TAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTC 301
```

```
Db    674 TGGAATCTCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCC 733

Qy    302 TAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAA 361

Db    734 TGACTTACCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAG 793

Qy    362 ATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACT 421

Db    794 GCGCCTTTACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACT 853

Qy    422 GGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCC 481

Db    854 GGAACCCTGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCC 913

Qy    482 ACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATAT 541

Db    914 ATACCTACGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCT 973

Qy    542 CTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAA 601

Db    974 CCAGTTGCAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGA 1033

Qy    602 AGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAA 661

Db   1034 AACTTTACAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAA 1093

Qy    662 ACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCA 721

Db   1094 ACTACGAGAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAA 1153

Qy    722 GTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACT 781

Db   1154 GAGAGCTTGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACT 1213

Qy    782 TTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCC 841

Db   1214 TCAACTGGAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTC 1273

Qy    842 TGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAAT 901

Db   1274 TTGCTTGTGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAAT 1333

Qy    902 GGTTTTGGAATTAA 915

Db   1334 GGTTTTGGAACTAA 1347
```

```
RESULT 10
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN
```

Query Match 64.4%; Score 589.2; DB 5; Length 1080;
Best Local Similarity 77.8%; Pred. No. 4.8e-146;
Matches 711; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

```
Qy      2 CAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGA 61

Db    167 CTAAACAGAATTATTACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGA 226

Qy     62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAG 121

Db    227 CTTTTGAACTGAAATCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTG 286

Qy    122 ACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGG 181

Db    287 ACCGCGCACTCTACAACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGG 346

Qy    182 ATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATT 241

Db    347 ATCTCATTAACCTGCCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATC 406

Qy    242 TAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTC 301

Db    407 TGGAATCACCAACACATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTC 466

Qy    302 TAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAA 361

Db    467 TGACTTACAGGCGTGACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAA 526

Qy    362 ATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACT 421

Db    527 AACCATTTGACTTTGAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACT 586

Qy    422 GGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCC 481

Db    587 GGAACCCAGAACATGCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTA 646

Qy    482 ACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATAT 541

Db    647 TAACTTATGGGCAAGCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTG 706

Qy    542 CTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAA 601

Db    707 CATCTTGCAAATTTTATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAA 766

Qy    602 AGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAA 661

Db    767 AACTTTATAATGCATTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAA 826

Qy    662 ACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCA 721

Db    827 ACTATGAGAACTACATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAA 886
```

```
Qy     722 GTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACT 781
Db     887 GAGAATTAGCTGACCATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACT 946

Qy     782 TTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCC 841
Db     947 ACAACTGGAGGAAACATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCT 1006

Qy     842 TGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAAT 901
Db    1007 TAGCGTGTGACCATGTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGT 1066

Qy     902 GGTTTTGGAATTAA 915
Db    1067 GGTTCTGGAACTAA 1080
```

RESULT 11
AB023627
LOCUS AB023627 1567 bp DNA linear VRT 02-APR-1999
DEFINITION Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete cds.
ACCESSION AB023627
VERSION AB023627.1 GI:4587295
KEYWORDS alpha(1,3)fucosyltransferase.
SOURCE Danio rerio DNA.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (sites)
AUTHORS Kageyama,N., Natsuka,S. and Hase,S.
TITLE Molecular cloning and characterization of two zebrafish alpha(1,3)fucosyltransferase genes developmentally regulated in embryogenesis
JOURNAL J. Biochem. (1999) In press
REFERENCE 2 (bases 1 to 1567)
AUTHORS Kageyama,N., Natsuka,S. and Hase,S.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034, Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381, Fax:81-6-6850-5383)
FEATURES Location/Qualifiers
source 1. .1567
/organism="Danio rerio"
/db_xref="taxon:7955"
gene 177. .1352
/gene="zFT1"
CDS 177. .1352
/gene="zFT1"
/codon_start=1
/product="alpha(1,3)fucosyltransferase"
/protein_id="BAA76706.1"
/db_xref="GI:4587296"
/translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT 499 a 334 c 293 g 441 t
ORIGIN

Query Match 29.3%; Score 268.2; DB 5; Length 1567;
Best Local Similarity 57.8%; Pred. No. 1.3e-60;
Matches 521; Conservative 0; Mismatches 368; Indels 12; Gaps 2;

```
Qy      18 TAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATC 77
Db     461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy      78 CTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAA 137
Db     521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy     138 CAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACC 197
Db     581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGGAT---------TTGCC 631

Qy     198 TCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCA 257
Db     632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy     258 CACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGA 317
Db     692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy     318 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTT 374
Db     752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy     375 TGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 434
Db     812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy     435 TGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCA 494
Db     872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy     495 AGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATT 554
Db     932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy     555 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 614
Db     992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy     615 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 674
Db    1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy     675 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAA 734
Db    1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy     735 ATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAA 794
Db    1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy     795 GGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCA 854
Db    1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy     855 TGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 914
Db    1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy     915 A 915
Db    1352 A 1352
```

RESULT 12
AB023628
LOCUS AB023628 1832 bp DNA linear VRT 02-APR-1999
DEFINITION Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete cds.
ACCESSION AB023628
VERSION AB023628.1 GI:4587297
KEYWORDS alpha(1,3)fucosyltransferase.
SOURCE Danio rerio DNA.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

```
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press .
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match             24.9%;  Score 228.2;  DB 5;  Length 1832;
  Best Local Similarity   55.0%;  Pred. No. 5.6e-50;
  Matches  496;  Conservative    0;  Mismatches  393;  Indels   12;  Gaps    2;

Qy      18 TAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATC 77
           | | || |  | | |  || | ||| |    || ||  |     | |||||||| | |
Db     619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy      78 CTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAA 137
            ||        ||  || ||||| || || ||  | |||  ||| |     ||  | |
Db     679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy     138 CAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACC 197
           ||| | ||||  ||  || ||||||| ||||| || |||   |||          || ||
Db     739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGCACAGAT---------TTTCC 789

Qy     198 TCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCA 257
           || |   | | |||   || || || ||||| |||  |||| | ||||| || || ||
Db     790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy     258 CACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGA 317
             |   |||||      | |  |  |   ||| ||  |||| |  || |||   | | ||
Db     850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy     318 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA---AATCCCTTTGTGTT 374
           |||||||||  | |||||||||  ||     |    |||     | |       |||
Db     910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy     375 TGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 434
             | || |  ||||| ||  | || ||||| | || |||||||   |      | ||
Db     970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy     435 TGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCA 494
               |  | |  | | ||||| |    ||    ||| |  |   |||||||||
Db    1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy     495 AGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATT 554
           |||    |     | || | | |  ||          | |||| | |||||||||
Db    1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy     555 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 614
           || || || ||||| ||||| ||||||| ||| ||  |||| || ||||| | |||| |
Db    1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy     615 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 674
           ||       ||| |||| ||||| ||  | || ||  | ||  | |||||||  | |
Db    1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy     675 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAA 734
           || |||   ||  | ||||| |||||||| |||||     ||||||  ||  | ||  |
Db    1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy     735 ATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAA 794
           | || |      | ||| ||||| |  |   ||||      |  || |||||||| ||
Db    1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy     795 GGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCA 854
           || |  |  |||     ||  |||| ||     |||| | ||   |   || ||  |
Db    1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy     855 TGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 914
           | | |||||| || |  | |||   | |||    || |  |    ||||| |||  |||
Db    1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy     915 A 915
           |
Db    1510 A 1510


RESULT 13
AF288369
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN
```

```
  Query Match             24.1%;  Score 220.6;  DB 5;  Length 1264;
  Best Local Similarity   57.2%;  Pred. No. 5.8e-48;
  Matches  441;  Conservative    0;  Mismatches  324;  Indels    6;  Gaps    2;

Qy      28 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 87
           ||||| ||| | |||  |||||| || || ||   |||| |  | |   |||| ||
Db     175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy      88 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 147
             | ||  | | ||     |||||  |||| ||  |||||   | || | | ||  |  ||
Db     235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy     148 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 207
           || ||  | || || || || ||  |    |||||       |   |  | | |    ||
Db     295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy     208 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 267
           ||||  || |  |||   ||||| ||| | ||  | ||||| || | |||  | ||  |
Db     355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy     268 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 327
             ||| ||| || ||| | ||||||||||||| |  | || ||  | || || || |||
Db     415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy     328 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 387
             || |||||     ||  | ||  ||   | |||      ||     | |||  |
Db     475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy     388 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 447
           |||   ||| ||||| |||||| || ||||| |||  ||         | ||||  |||
Db     535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy     448 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 507
           || ||| | ||||| |  ||    ||     |  |    || || |    |  |
Db     595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGCGGCAACAC---TTG 651

Qy     508 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 567
             | ||  |    | |  ||     |  ||| ||||    |  || || ||  |  | |||
Db     652 CCCCTGTCCATGGACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy     568 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTAC---AATGCATTTTTGGCT 624
           || ||||||   |||||||| |||||||| || |||||||     |||||  | |   ||
Db     712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy     625 GGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCT 684
           ||  |  | || ||||| ||||| ||| || | |||||||||||| | |   | ||  ||
Db     772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy     685 GATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAG 744
           || || ||||| || || || || ||| |  | | ||||  | ||  | || ||| ||
Db     832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy     745 GAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
           ||| |   ||  |||  ||  |||||   | |||||  | ||| | | |
Db     892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942


RESULT 14
AB035907
LOCUS       AB035907               1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN

  Query Match             22.6%;  Score 206.6;  DB 5;  Length 1092;
  Best Local Similarity   56.1%;  Pred. No. 3.1e-44;
  Matches  474;  Conservative    0;  Mismatches  359;  Indels   12;  Gaps    4;

Qy      19 AACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCC 78
           ||  ||   ||  |  |  | |||     |||||||||  |           |||
Db     199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy      79 TGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAAC 138
           |||||     | |||||| ||   ||||||||  |||||||  || |     | ||| |
Db     259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy     139 AAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCT 198
           |  |  | ||  |  || | |||||  | |||||    |  ||| | | | || | |||
Db     319 GAGGCAGACGCCATTATGTTCCACCACCGGGACATT---TCTGATTTCAGTGACCTTCCT 375

Qy     199 CAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCAC 258
           |      | |    ||  || || ||||| |||||||| || || || ||  | ||
Db     376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy     259 ACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGAT 318
           | |||     |   |  | |    | | ||||||  |||  |  | ||  |    ||
Db     436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy     319 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAA 378
           || ||||||    |||| || ||||||||     |   ||  || |          | |
Db     496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy     379 GTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 438
           ||||  |||||  ||| ||||| ||||||  | || |||||||||   || || |
Db     553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy     439 AGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCA 498
           |||||| ||||||||||||||||| || ||      | |||||| ||   || ||||
Db     613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy     499 TTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTAT 558
           |  ||      |  | || ||  | ||  |  |  ` ||| | ||    | |||||||||
Db     670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy     559 CTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCA 615
           |  | |||||  || ||  | ||||  ||||||||||| |||||||| |     || ||
```

```
Db    730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy    616 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 675

Db    790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy    676 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 735

Db    850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy    736 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795

Db    910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy    796 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 855

Db    970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy    856 GTAAA 860

Db   1030 GTCAA 1034


RESULT 15
GGU73678
LOCUS       GGU73678               2657 bp    DNA     linear   VRT 20-DEC-1996
DEFINITION  Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete
            cds.
ACCESSION   U73678
VERSION     U73678.1  GI:1657998
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N.,
            Schultz,T.L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
  TITLE     Molecular cloning and characterization of CFT1, a developmentally
            regulated avian alpha(1,3)-fucosyltransferase gene
  JOURNAL   J. Biol. Chem. 271 (51), 32960-32967 (1996)
  MEDLINE   97115837
REFERENCE   2  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M. and Larsen,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical
            Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES             Location/Qualifiers
     source          1. .2657
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            170. .1240
                     /gene="CFT1"
     CDS             170. .1240
                     /gene="CFT1"
                     /note="similar to mammalian fucosyltransferase FucTIV"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase"
                     /protein_id="AAC60060.1"
                     /db_xref="GI:1657999"
                     /translation="MELGPRWSPAARPGCPRRWRRRWALLGALLGAALALYVCVRELR
                     RRGSAAGRPEGEVTVLLWWEPFGRPWRPADCRRRYNITGCLLSADRGRYGEARAVLFH
                     HRDLALHGRQGLPRGPPPRPPRQRWVWMNFESPSHSPGLRGLAGLFNWTMSYRRDSDV
                     FVPYGYLYEPPSPRPFVLPRKSRLVAWVISNWNEEHARVRYYRQLKEHLPIDVYGARG
                     MALLEGSVVKTVSAYKFYLAFENSQHTDYITEKLWKNAFAASAVPVVLGPRRANYERF
                     IPADSFIHVDDFPSPRLLATYLKFLDKNKPSYRRYFAWRNKYEVHVTSFWDEHYCKVC
                     EAVRTAGNQLKTVQNLAGWFES"
BASE COUNT      581 a    712 c    818 g    546 t
ORIGIN

  Query Match           19.1%;  Score 175.2;  DB 5;  Length 2657;
  Best Local Similarity 52.9%;  Pred. No. 7.1e-36;
    Matches  453;  Conservative    0;  Mismatches  388;  Indels   15;  Gaps    3;

Qy     21 CGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTG 80

Db    334 CGAGGTGACGGTGCTGCTGTGGTGGGAGCCCTTCGGCCGCCCCTGGCGCCCGGCCGACTG 393

Qy     81 CCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAA 140

Db    394 CCGGAGGCGCTACAACATCACGGGCTGCCTCCTGAGCGCCGACCGCGGGCGGTACGGCGA 453

Qy    141 ATCCCATGCGGTCCTGATTCACCATAGAGACAT---------CAGCTGGGATCTGACTAA 191

Db    454 GGCTCGGGCCGTGCTCTTCCACCACCGCGACCTGGCGCTGCACGGCCGCCAGGGGCTGCC 513

Qy    192 CTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACC 251

Db    514 CCGCGGGCCCCCGCCGCGACCCCCGCGGCAGCGGTGGGTGTGGATGAACTTCGAGTCGCC 573

Qy    252 CACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCG 311

Db    574 CTCGCACTCCCCCGGGCTGCGGGGCCTGGCCGGGCTCTTCAACTGGACCATGTCGTACCG 633

Qy    312 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGT 371

Db    634 ACGGGACTCGGACGTCTTCGTGCCCTACGGGTACCTCTACGAGCCGCCGTCGCCGCGGCC 693

Qy    372 GTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGA 431

Db    694 CTTCGTGCTGCCTCGCAAAAGCCGGCTGGTGGCCTGGGTCATCAGCAACTGGAACGAGGA 753

Qy    432 GCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGG 491

Db    754 GCACGCCCGCGTGCGCTACTACCGGCAGCTGAAGGAGCACCTGCCCATCGACGTGTACGG 813

Qy    492 CCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAA 551

Db    814 ---GGCGCGCGGGATGGCGCTGCTGGAGGGCAGCGTGGTGAAGACGGTGTCGGCCTACAA 870

Qy    552 ATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT---A 608

Db    871 GTTCTACCTGGCCTTCGAGAACTCCCAGCACACGGACTACATCACCGAGAAGCTGTGGAA 930

Qy    609 CAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGA 668

Db    931 GAACGCCTTCGCCGCCAGCGCGGTGCCTGTTGTCCTCGGCCCCCGCAGGGCCAACTACGA 990

Qy    669 GAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTT 728

Db    991 GCGCTTCATCCCCGCCGACTCCTTCATCCACGTGGACGACTTCCCCAGCCCACGGCTGCT 1050

Qy    729 AGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTG 788

Db   1051 GGCCACCTACCTGAAGTTCCTCGATAAAAACAAGCCCAGCTACAGGAGGTATTTCGCCTG 1110

Qy    789 GAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATG 848

Db   1111 GCGGAACAAGTATGAAGTCCACGTCACGTCTTTCTGGGATGAGCATTACTGCAAGGTTTG 1170

Qy    849 CGACCATGTAAAAAGG 864

Db   1171 CGAGGCCGTGAGGACG 1186


Search completed: October  6, 2002, 04:12:59
Job time : 1338.59 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 102.682 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-3_COPY_280_1194
Perfect score: 915
Sequence: 1 acaaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	915	100.0	2036	21	AAZ92672	Murine alpha-1,3-f
2	915	100.0	2170	21	AAZ92645	Murine alpha-1,3-f
3	791.8	86.5	2676	21	AAZ92647	Human alpha-1,3-fu
4	791.8	86.5	2822	21	AAZ92646	Human alpha-1,3-fu
5	129	14.1	1814	18	AAT59505	Murine myeloid-lin
6	128.6	14.1	3594	18	AAT85219	Mouse alpha-fucosy
7	128.2	14.0	1086	15	AAQ56912	Lewis blood group
8	128.2	14.0	2042	18	AAT80111	Fucosyltransferase
9	128.2	14.0	2043	12	AAQ13330	Glycosyltransferas
10	128.2	14.0	2043	15	AAQ56906	DNA encoding a gly
11	128.2	14.0	2043	18	AAT76769	Human alpha 1,3/4
12	128.2	14.0	2043	18	AAT61675	Human alpha(1,3/1,
13	126.6	13.8	1316	15	AAQ56911	DNA encoding a gly
14	126.6	13.8	1316	18	AAT61679	Human alpha(1,3)-f
15	124.6	13.6	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
16	124.6	13.6	1654	18	AAT61680	Human alpha(1,3)-f
17	123	13.4	1126	21	AAF21135	Human low adenosin
18	123	13.4	1126	21	AAA35013	Human adenosine re
19	123	13.4	6944	21	AAF21137	Human low adenosin
20	123	13.4	6944	21	AAA35015	Human adenosine re
21	122.8	13.4	1701	15	AAQ77732	Human alpha-1,3-fu
22	122.8	13.4	1701	21	AAF21136	Human low adenosin
23	122.8	13.4	1701	21	AAA35014	Human adenosine re
24	120	13.1	1256	21	AAF21134	Human low adenosin
25	120	13.1	1256	21	AAA35012	Human adenosine re
26	120	13.1	1400	13	AAQ31436	Encodes a HeLa cel
27	120	13.1	1488	15	AAQ56910	DNA encoding a gly
28	120	13.1	2175	12	AAQ14382	Clone 7.2 encoding
29	120	13.1	2175	24	AAS17082	Human cDNA clone 7
30	120	13.1	2861	12	AAQ14383	Clone 1 encoding 1
31	120	13.1	2861	21	AAF21133	Human low adenosin
32	120	13.1	2861	21	AAA35011	Human adenosine re
33	120	13.1	2861	24	AAS17083	Human cDNA clone 1
34	120	13.1	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
35	120	13.1	3647	15	AAQ56909	DNA encoding a gly
36	120	13.1	3647	18	AAT61678	Human alpha(1,3)-f
37	118.4	12.9	2175	11	AAQ06691	CDX, a MILA involv
38	116.8	12.8	2134	18	AAT59506	Human myeloid deri
39	116	12.7	795	17	AAT13798	Alpha-(1,3/1/4)-fu
40	64.6	7.1	1578	23	ABL11867	Drosophila melanog
c 41	64.6	7.1	4229	23	ABL11866	Drosophila melanog
42	53	5.8	1353	21	AAZ51685	C. elegans alpha-1
43	43.6	4.8	2198	21	AAA97937	Mung bean alpha 1,
44	40.8	4.5	398	22	AAF64428	Novel human polynu
45	40.8	4.5	676	22	AAS41344	cDNA encoding nove

ALIGNMENTS

RESULT 1
AAZ92672
ID AAZ92672 standard; cDNA; 2036 BP.
XX
AC AAZ92672;
XX
DT 05-JUN-2000 (first entry)
XX
DE Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; murine; mouse; ss.
XX
OS Mus sp.
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x

```
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Example 8; Page 162-164; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

  Query Match          100.0%;  Score 915;  DB 21;  Length 2036;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-261;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 233

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 293

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     294 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 353

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 413

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     414 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 473

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     474 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 533

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     534 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 593

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     594 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 653

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     654 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 713

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     714 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 773

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     774 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 833

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     834 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 893

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     894 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 953

Qy     781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     954 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1013

Qy     841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1073

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    1074 TGGTTTTGGAATTAA 1088


RESULT 2
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag= a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
```

```
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match             100.0%;  Score 915;  DB 21;  Length 2170;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-261;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 339

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     340 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 399

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     400 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 459

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 519

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 579

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 639

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 699

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 759

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 819

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 879

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 939

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 999

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 1059

Qy     781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1119

Qy     841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1179

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    1180 TGGTTTTGGAATTAA 1194


RESULT 3
AAZ92647
ID   AAZ92647 standard; cDNA; 2676 BP.
XX
AC   AAZ92647;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             295..1374
FT                   /*tag=  a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
```

```
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match           86.5%;  Score 791.8;  DB 21;  Length 2676;
  Best Local Similarity  91.6%;  Pred. No. 3.6e-224;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy      1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          || |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db    460 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 519

Qy     61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db    520 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 579

Qy    121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
          ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db    580 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 639

Qy    181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
          |||||||| || |||||||||||| || |||||||||||| ||||||||||||||||||||
Db    640 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 699

Qy    241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
          || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db    700 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 759

Qy    301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
          || ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    760 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 819

Qy    361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
          |||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db    820 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 879

Qy    421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db    880 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 939

Qy    481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
          || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db    940 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 999

Qy    541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
          ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db   1000 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 1059

Qy    601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
          ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db   1060 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 1119

Qy    661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
          |||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db   1120 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 1179

Qy    721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
          |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db   1180 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1239

Qy    781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db   1240 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1299

Qy    841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1300 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1359

Qy    901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db   1360 TGGTTTTGGAATTAA 1374


RESULT 4
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag=  a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 134-142; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
```

```
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

  Query Match           86.5%;  Score 791.8;  DB 21;  Length 2822;
  Best Local Similarity  91.6%;  Pred. No. 3.6e-224;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          || ||||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db    1454 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 1513

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
           |||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db    1514 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 1573

Qy     121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 180
           ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db    1574 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 1633

Qy     181 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db    1634 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 1693

Qy     241 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db    1694 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 1753

Qy     301 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1754 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 1813

Qy     361 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||
Db    1814 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 1873

Qy     421 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 480
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db    1874 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 1933

Qy     481 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 540
           || ||||| || |||||||| |||||||| || || |||||||| ||||||| ||||||
Db    1934 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 1993

Qy     541 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 600
           ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db    1994 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 2053

Qy     601 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 660
           ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db    2054 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 2113

Qy     661 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 720
           |||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Db    2114 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 2173

Qy     721 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 780
           |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db    2174 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 2233

Qy     781 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 840
           |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db    2234 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 2293

Qy     841 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    2294 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 2353

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db    2354 TGGTTTTGGAATTAA 2368


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag=  a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
```

```
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             14.1%;  Score 129;  DB 18;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 9e-28;
  Matches  335;  Conservative    0;  Mismatches  300;  Indels    6;  Gaps    2;

Qy     104 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 163
           | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     164 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 223
           |  | ||  |          |   ||  | |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 283
             ||| | |||       | || || |||| | | ||||          | || |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           | | ||||||  |   || |  ||||| ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 403
             ||||  |     | |    |||      |    | | ||   |||     | | |||
Db     788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 463
           ||||||  | || || |   |       |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 523
           |      | |  |  |  |     |  || |  ||   ||||      | |   || |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 583
           ||||  |  |||||   |  |     |      || || ||  | ||||| ||||||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 640
             || |||||||| || ||| |||     ||||||  |   |||||||  | ||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 700
           |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTG 741
           |||| || || | |||| || ||||  | ||   |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag=  a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag=  b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag=  c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag=  d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag=  e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag=  f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag=  g
FT                   /number=  3
FT   exon            2067..3552
FT                   /*tag=  f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             14.1%;  Score 128.6;  DB 18;  Length 3594;
  Best Local Similarity   52.5%;  Pred. No. 1.6e-27;
  Matches  331;  Conservative    0;  Mismatches  294;  Indels    6;  Gaps    2;

Qy     104 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 163
           | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336
```

```
Qy     164 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 223
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     224 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 283
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     284 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     344 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 403
Db    2517 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 463
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     464 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 523
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     524 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 583
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy     584 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 640
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy     641 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 700
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy     701 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 731
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904


RESULT 7
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
DE   Lewis blood group fucosyltransferase DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding a Lewis blood group fucosyl
CC   transferase (Fuc-TIII).  The enzyme produced by the DNA may be
CC   non glycosylated.  This prevents premature loss of enzyme activity.
CC   It can also be used in in vitro reactions to modify cell surface
CC   oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-11.
XX
SQ   Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

  Query Match             14.0%;  Score 128.2;  DB 15;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 1.2e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
Db     271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy     166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
Db     331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy     223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
Db     391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy     283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
Db     451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy     343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
Db     511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy     394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
Db     571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy     454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
Db     631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy     514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
Db     688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy     574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy     631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy     691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy     751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 8
AAT80111
ID   AAT80111 standard; cDNA; 2042 BP.
XX
AC   AAT80111;
XX
DT   01-DEC-1997  (first entry)
XX
DE   Fucosyltransferase III coding sequence.
```

```
XX
KW   P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW   calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW   thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW   inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW   disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW   adult respiratory distress syndrome; tumour metastasis;
KW   fucosyltransferase III; acetylglucosaminyltransferase;
KW   autoimmune disorder; inflammatory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9706176-A2.
XX
PD   20-FEB-1997.
XX
PF   02-AUG-1996;   96WO-US12820.
XX
PR   17-MAY-1996;   96US-0649802.
PR   03-AUG-1995;   95US-0510920.
PR   15-MAY-1996;   96US-0017794.
XX
PA   (OKLA ) UNIV OKLAHOMA STATE.
XX
PI   Cummings RD,  Mcever RP,  Moore KL;
XX
DR   WPI; 1997-154206/14.
XX
PT   New O-glycan(s) derived from P-selectin glyco:protein ligand
PT   (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT   for treating inflammatory or autoimmune disorders or tumours
XX
PS   Disclosure; Page 80-81; 96pp; English.
XX
CC   This sequence represents the coding sequence for the human
CC   fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC   acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC   expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC   mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC   protein expressed on the surfaces of activated platelets and endothelium
CC   in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC   high affinity P-selectin ligand produced by leukocytes. Binding of
CC   P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC   treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC   highly extended extracellular domain, which is a feature of mucin-like
CC   proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC   the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC   O-glycans, and O-glycanated products of the invention. The products are
CC   used for inhibiting the binding of P-selectin and other selectins to
CC   PSGL-1. They can be used for the treatment of inflammation, injury
CC   resulting from ischaemia and reperfusion, bacterial sepsis and
CC   disseminated intravascular coagulation, adult respiratory distress
CC   syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC   other autoimmune or inflammatory disorders. The products can also be used
CC   in the production of antibodies and in diagnostic applications.
XX
SQ   Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

  Query Match             14.0%;  Score 128.2;  DB 18;  Length 2042;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          ||||| |||| | |||||||    |||||    |  |  | ||||| | | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
           | || |||| |   |   |  | |   | | || |      | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
          |||||  ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |    |    ||          | |  |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || |||||  |||||  |||   ||||||||| || ||    |||||||     || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
            | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  | |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |         ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 9
AAQ13330
ID   AAQ13330 standard; DNA; 2043 BP.
XX
AC   AAQ13330;
XX
DT   07-NOV-1991  (first entry)
XX
DE   Glycosyltransferase gene in clone pCDM7-alpha(1,3/1,4)FT.
XX
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..1158
FT                   /*tag=  a
FT   polyA_site      2023..2034
FT                   /*tag=  b
FT   polyA_signal    1926..2003
FT                   /*tag=  c
FT   repeat_region   1183..1547
FT                   /*tag=  d
FT                   /note= "20 copies of a degenerate 16bp sequence"
FT   repeat_region   1885..1999
FT                   /*tag=  e
FT                   /note= "similar to human Alu sequences"
XX
PN   WO9112340-A.
XX
PD   22-AUG-1991.
XX
PF   14-FEB-1991;   91WO-US00899.
XX
PR   12-DEC-1990;   90US-0627621.
```

```
PR   14-FEB-1990;   90US-0479858.
PR   14-FEB-1990;   90US-0480133.
XX
PA   (UNMI ) UNIV OF MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1991-267151/36.
DR   P-PSDB; AAR13749.
XX
PT   Isolation of gene conveying post-translational characteristic -
PT   e.g. the presence of soluble or membrane bound oligo or
PT   polysaccharide or glycosyltransferase.
XX
PS   Claim 11; Fig 1; 155pp; English.
XX
CC   The DNA encodes a protein sequence capable of functioning as a
CC   GDP- Fuc:[beta-D-Gal(1,4/1,3)]-D-GlcNAc(/Glc)alpha (1,3/1,4)-
CC   fucosyltransferase. The sequence coded by nucleotides 127 to 1083
CC   represents the functional protein. The DNA sequence can be used to
CC   construct animal cell lines with respect to posttranslational
CC   modification of the oligosaccharides on cell-surface, intracellular
CC   or secreted proteins or lipids by sub-terminal alpha(1,3) and alpha
CC   (1,4) fucose residues that represent the products of this enzyme.
CC   These can be used for the prodn. of diagnostics and therapeutics.
CC   See also AAQ13331-Q13333.
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match             14.0%;  Score 128.2;  DB 12;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          ||||| |||| | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
            | || |||| |   |   |  | |   | | || |      | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
            ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||         | |  |  |  ||   || || |   |   |    |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
            | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || |||||||| || ||||| |     || || |  |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    |||||  |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 10
AAQ56906
ID   AAQ56906 standard; DNA; 2043 BP.
XX
AC   AAQ56906;
XX
DT   26-JUL-1994  (first entry)
XX
DE   DNA encoding a glycosyltransferase.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             72..1158
FT                   /*tag= a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45934.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 1; 249pp; English.
XX
CC   The sequence is that encoding human glycosyl transferase.  The enzyme
CC   produced by the DNA may be non glycosylated.  This prevents premature
CC   loss of enzyme activity.  It can also be used in in vitro reactions to
CC   modify cell surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-12.
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;

  Query Match             14.0%;  Score 128.2;  DB 15;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          ||||| |||| | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
            | || |||| |   |   |  | |   | | || |      | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
            ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522
```

```
Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 11
AAT76769
ID   AAT76769 standard; cDNA; 2043 BP.
XX
AC   AAT76769;
XX
DT   15-SEP-1997  (first entry)
XX
DE   Human alpha 1,3/4 fucosyltransferase cDNA.
XX
KW   GDP-L-fucose:beta-D-N-acetylglucosamide 3/4-alpha-L-fucosyltransferase;
KW   human; alpha 1,2 FT; transgene; transgenic mouse; animal model;
KW   intestinal adhesion; Helicobacter pylori infection; stomach;
KW   small intestine; gut; epithelial cell; surface receptor;
KW   carbohydrate antigen; gastritis; peptic ulcer; neoplasia;
KW   gastric adenocarcinoma; Lewis antigen; fucosylation; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             104..1201
FT                   /*tag=  a
XX
PN   US5625124-A.
XX
PD   29-APR-1997.
XX
PF   11-JUL-1994;   94US-0273411.
XX
PR   11-JUL-1994;   94US-0273411.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Falk P,  Gordon JI;
XX
DR   WPI; 1997-258275/23.
DR   P-PSDB; AAW23806.
XX
PT   Animal model for Helicobacter pylori infection - comprising
PT   transgenic mouse expressing human enzyme promoting intestinal
PT   adhesion
XX
PS   Example 1; Columns 23-26; 24pp; English.
XX
CC   A claimed transgenic mouse expresses, in its intestinal epithelial
CC   cells, the enzyme human GDP-L-fucose:beta-D-galactoside 2-alpha-L-
CC   fucosyltransferase (also called alpha 1,2 FT) or human GDP-L-fucose:
CC   beta-D-N-acetylglucosamide 3/4-alpha-L-fucosyltransferase (also
CC   called alpha 1,3/4 FT). The enzyme is expressed under the
CC   control of a gut epithelial cell-specific promoter and Helicobacter
CC   pylori adheres to the transgenic cells. The transgenic mouse and
CC   intestinal epithelial cells from it are useful as models for screening
CC   compounds for the ability to inhibit adhesion of H. pylori to gut
CC   epithelial cells. The present sequence encodes human alpha 1,3/4 FT
CC   and was published by Kukowska-Latallo et al., Genes Dev., 4:
CC   1288-1303 (1990).
XX
SQ   Sequence 2043 BP; 413 A; 647 C; 575 G; 408 T; 0 other;
```

```
  Query Match             14.0%;  Score 128.2;  DB 18;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
```

```
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 12
AAT61675
ID   AAT61675 standard; cDNA; 2043 BP.
XX
AC   AAT61675;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3/1,4)-fucosyltransferase (Lewis enzyme) cDNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TIII;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..1158
FT                   /*tag=  a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13638.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 3; Page 268-269; 329pp; English.
XX
CC   A cDNA clone (AAT61675) codes for an enzyme (AAW13638) that functions as
CC   a GDP-Fuc:(beta-D-Gal(1,4/1,3))-D-GlcNAc(/Glc) alpha(1,3/1,4)-
CC   fucosyltransferase (Fuc-TIII) which, when expressed by the cloned
CC   cDNA sequence, functions within mammalian cells to generate de novo
CC   expression of specific SSEA-1 (Lewis x), sialyl-Lewis x, Lewis a and
CC   sialyl-Lewis a structures.  The clone was identified by transfecting
CC   COS-1 cells with genomic DNA from human A431 cells, and isolating a
CC   cloned cDNA that detd. expression of surface-localised SSEA-1
CC   structures.  The cDNA can be used to construct animal cell lines
CC   with specific capabilities with respect to post-translational
CC   modification of the oligosaccharides of expressed proteins or
CC   lipids, to produce recombinant Fuc-TIII for use in oligosaccharide
CC   prodn., to genotype individuals for the Lewis locus, and in the
CC   prodn. of chimeric fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 2043 BP; 413 A; 646 C; 576 G; 408 T; 0 other;

  Query Match             14.0%;  Score 128.2;  DB 18;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-27;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
           |||||  ||||  | ||||||    |||||    |  |  |  |||||  |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
             | || ||||  |   |   |  | |   | | || |      | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
               ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||

Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
             |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
           |   ||  |  | |  |     |    ||           | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
           |    ||          | |  |  |  ||  || ||  |   |   |    |  | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
             | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
           ||  | |||   || ||||||||| || ||||| |     || ||  |  |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
           || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 13
AAQ56911
ID   AAQ56911 standard; DNA; 1316 BP.
XX
AC   AAQ56911;
XX
DT   26-JUL-1994  (first entry)
XX
DE   DNA encoding a glycosyltransferase.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45939.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
```

```
PS   Disclosure; Fig 6; 249pp; English.
XX
CC
CC   The sequence is that encoding human glycosyl transferase. The enzyme
CC   produced by the DNA may be non glycosylated. This prevents premature
CC   loss of enzyme activity. It can also be used in in vitro reactions to
CC   modify cell surface oligosaccharide mols. e.g. blood group determinants.
XX   See also AAQ56905-12.
SQ
     Sequence 1316 BP; 236 A; 456 C; 372 G; 252 T; 0 other;

  Query Match                13.8%;  Score 126.6;  DB 15;  Length 1316;
  Best Local Similarity      52.9%;  Pred. No. 3.9e-27;
  Matches  373;  Conservative    0;  Mismatches  314;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          ||| | ||||  | |||  |  |||||  | | | | | |||||| |  | |||||
Db    430 TGCAACATCACTGCCGACTCCAGTGTGTACCCACAGGCAGACGCGGTCATCGTGCACCAC 489

Qy    166 AGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          || |||| |  | |  | |||   |||| | | || |  | |  |||||| |  |||
Db    490 TGGGATATCATGTACAACCCCAGTGCCAACCTCCCGCCCCCCACCAGGCCGCAGGGGCAG 549

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
          |||||| ||| |  | |||| ||  ||  ||  |  |  | | | | | | | | |||
Db    550 CGCTGGATCTGGTTCAGCATGGAGTCCCCCAGCAACTGCCGGCACCTGGAAGCCCTGGAC 609

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          | ||||| || ||  | | | ||| || || || || ||| |  | | | | ||| |||
Db    610 GGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 669

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          || | | | | |  |  |  ||  |||         |  | | | |  | | | ||||
Db    670 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 729

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
          || |||||  |||||  |||  |||||||| || |  | | | | | | | | | |||
Db    730 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCGGCCAGGGTGCGCTACTAC 789

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    790 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 846

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    847 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTATCTGGCCTTCGAGAAC 906

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  |||  |  |||||||| || |||||  |||   |  | | | | | | | | | |
Db    907 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 966

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          ||  ||  ||  |||||  |  |||  |||| |||| | | | | | | | | | | |
Db    967 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCGCCCGACGCC 1026

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          |||||  ||  |||||  ||  |  |  |||  |  |  | | | | | | | | | |||
Db   1027 TTCATCCACGTGGATGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 1086

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  ||  |  |||||  |||| |  |||  |  |  |  |
Db   1087 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1131

RESULT 14
AAT61679
ID   AAT61679 standard; DNA; 1316 BP.
XX
AC   AAT61679;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TV) DNA.
XX
KW   Alpha(1,3)-fucosyltransferase; Lewis enzyme; Fuc-TV;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key
FT   CDS                 Location/Qualifiers
FT                       121..1245
XX                       /*tag=  a
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;  96WO-US13816.
XX
PR   08-SEP-1995;  95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ, Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13642.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 5; Page 287-288; 329pp; English.
XX
CC   A DNA clone (AAT61679) codes for an enzyme (AAW13642) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TV) which can utilise neutral type II oligosaccharide
CC   precursors but which cannot efficiently utilise type I
CC   glycococonjugates. It was identified in a human lambda phage
CC   genomic DNA library using a probe obtd. from Lewis enzyme cDNA (see
CC   also AAT61675). The Fuc-TV enzyme, when expressed by the cloned DNA
CC   sequence, functions within mammalian cells to generate de novo
CC   expression of specific cell surface glycoconjugate structures that
CC   are recognised by antibodies against SSEA-1, Lewis x or sialyl
XX   Lewis x.
SQ
     Sequence 1316 BP; 236 A; 456 C; 372 G; 252 T; 0 other;

  Query Match                13.8%;  Score 126.6;  DB 18;  Length 1316;
  Best Local Similarity      52.9%;  Pred. No. 3.9e-27;
  Matches  373;  Conservative    0;  Mismatches  314;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          ||| | ||||  | |||  |  |||||  | | | | | |||||| |  | |||||
Db    430 TGCAACATCACTGCCGACTCCAGTGTGTACCCACAGGCAGACGCGGTCATCGTGCACCAC 489

Qy    166 AGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          || |||| |  | |  | |||   |||| | | || |  | |  |||||| |  |||
Db    490 TGGGATATCATGTACAACCCCAGTGCCAACCTCCCGCCCCCCACCAGGCCGCAGGGGCAG 549

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
          |||||| ||| |  | |||| ||  ||  ||  |  |  | | | | | | | | |||
Db    550 CGCTGGATCTGGTTCAGCATGGAGTCCCCCAGCAACTGCCGGCACCTGGAAGCCCTGGAC 609

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          | ||||| || ||  | | | ||| || || || || ||| |  | | | | ||| |||
Db    610 GGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 669

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          || | | | | |  |  |  ||  |||         |  | | | |  | | | ||||
Db    670 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 729

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
          || |||||  |||||  |||  |||||||| || |  | | | | | | | | | |||
Db    730 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCGGCCAGGGTGCGCTACTAC 789

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
```

```
Db    790 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 846

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
          | |    .|   ||||    ||  | ||     |  || || |||||  | || || |||
Db    847 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTATCTGGCCTTCGAGAAC 906

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    907 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 966

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    ||||| |||| |   | ||  | ||  |
Db    967 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCGCCCGACGCC 1026

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    |||| ||  |.||    || ||| |||| |
Db   1027 TTCATCCACGTGGATGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 1086

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1087 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1131


RESULT 15
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             13.6%;  Score 124.6;  DB 15;  Length 1654;
  Best Local Similarity   52.1%;  Pred. No. 1.7e-26;
  Matches  408;  Conservative    0;  Mismatches  354;  Indels   21;  Gaps    5;

Qy     31 ATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
          || ||| | |||   ||||| |||    |  || | |  ||  |   |||| ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     91 TTCAAT---ATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 147
            |   |   |      |  ||| |  ||||  | |||||    ||||      |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    148 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA---CTAACTTACCTCAGCAG 204
          ||||||  |  | |||||  ||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    205 GCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCC 264
          | ||||  |     |||  |||||| ||| | |   | |||||  || |  |||  |
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    265 CAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGAT 324
          ||   ||  | ||| ||    |  ||||| || ||  |  | ||  ||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    325 ATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTT 375
          |||      ||| || ||||   ||  |  | |  |  |    ||          | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    376 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
          |  |  |   ||||    || |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    436 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 495
          ||||||||    || ||| |    ||          | |  |  |  ||   || ||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    496 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 555
            | |     |  | ||  | |   ||   ||||    ||   | ||     |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    556 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTAC---AAT 612
          |||||  | || || |||||  | |||   || |||||||| || |||||  |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    613 GCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAAT 672
          ||  |   |||  |  | || || || || ||||| ||    ||    |||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    673 TATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCA 732
          |   | ||| | ||  | ||||| || |||| || ||  |    |||  ||  | ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    733 AAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGA 792
            || ||| ||||  | ||||  || |       ||||| || ||||||  |||| |
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    793 AAG 795
          ||
Db   1095 GAG 1097


Search completed: October  5, 2002, 23:10:04
Job time : 121.682 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 766.105 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-3_COPY_280_1194
Perfect score: 915
Sequence: 1 acaaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	661.8	72.3	697	9	BB397507	BB397507 BB397507
2	580.8	63.5	607	10	BI102760	BI102760 602888338
3	368	40.2	649	9	AV338075	AV338075 AV338075
4	366	40.0	716	9	AV725071	AV725071 AV725071
c 5	291.2	31.8	1101	12	CNS059ST	AL327638 Tetraodon
c 6	272.2	29.7	579	10	BM021081	BM021081 ie75f03.x
c 7	179.6	19.6	1100	12	CNS0579Z	AL324368 Tetraodon
8	175.6	19.2	844	12	CNS01V8V	AL168808 Tetraodon
c 9	171.4	18.7	570	10	BM082484	BM082484 fu25e08.x
10	170.8	18.7	648	9	BB595164	BB595164 BB595164
c 11	164.8	18.0	887	12	CNS02Q0S	AL208693 Tetraodon
c 12	156.8	17.1	1016	12	CNS0429P	AL271222 Tetraodon
13	152.2	16.6	199	9	BB398103	BB398103 BB398103
14	150	16.4	625	9	BB649967	BB649967 BB649967
c 15	148.4	16.2	666	10	BJ050300	BJ050300 BJ050300
16	145.8	15.9	651	10	BJ036855	BJ036855 BJ036855
17	142	15.5	640	10	BJ073227	BJ073227 BJ073227
c 18	140.4	15.3	711	12	CNS034XF	AL228012 Tetraodon
19	140.4	15.3	767	9	BB640911	BB640911 BB640911
c 20	133.4	14.6	501	12	B41073	B41073 HS-1052-B2-
c 21	118.6	13.0	674	10	BE605503	BE605503 fl10h04.x
22	115.6	12.6	971	12	CNS047IV	AL278032 Tetraodon
c 23	110.8	12.1	1023	12	CNS02ANG	AL188773 Tetraodon
c 24	110.2	12.0	615	10	BI326811	BI326811 STR00078
25	106.2	11.6	487	10	BE863396	BE863396 UI-M-BH0-
26	106.2	11.6	720	10	BF180380	BF180380 601804564
c 27	105.8	11.6	702	10	BJ043578	BJ043578 BJ043578
28	90.8	9.9	628	9	BB651409	BB651409 BB651409
29	90	9.8	863	12	CNS02ROF	AL210840 Tetraodon
30	88.4	9.7	551	10	BJ035701	BJ035701 BJ035701
c 31	87.8	9.6	566	10	BJ090565	BJ090565 BJ090565
32	85.4	9.3	601	10	BJ036679	BJ036679 BJ036679
33	85.4	9.3	605	10	BJ028804	BJ028804 BJ028804
c 34	84.8	9.3	776	12	CNS03RIO	AL257289 Tetraodon
c 35	83.8	9.2	1056	12	CNS01VL7	AL169252 Tetraodon
36	82.6	9.0	638	10	BJ035697	BJ035697 BJ035697
37	81.8	8.9	655	10	BG538740	BG538740 602566977
c 38	81.4	8.9	878	12	CNS02V1D	AL215194 Tetraodon
39	80.8	8.8	304	9	AU081478	AU081478 AU081478
40	80.8	8.8	1057	12	CNS03ADC	AL235065 Tetraodon
41	74.8	8.2	838	10	BG432556	BG432556 602500710
42	74.6	8.2	590	10	BM082755	BM082755 fu25e08.y
c 43	73.2	8.0	881	12	CNS03732	AL230807 Tetraodon
44	71.4	7.8	556	9	AU052985	AU052985 AU052985
45	71.2	7.8	505	10	BJ029303	BJ029303 BJ029303

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN

  Query Match             72.3%;  Score 661.8;  DB 9;  Length 697;
  Best Local Similarity   99.6%;  Pred. No. 8.1e-171;
  Matches  674;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy    239 ATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGA 298
          | ||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db      4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy    299 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy    359 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy    419 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy    479 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy    539 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy    599 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy    659 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy    719 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy    779 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy    839 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy    899 AATGGTTTTGGAATTAA 915
          |||||||||||||||||
Db    663 AATGGTTTTGGAATTAA 679
```

```
RESULT 2
BI102760
LOCUS       BI102760                607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: 1  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN

  Query Match             63.5%;  Score 580.8;  DB 10;  Length 607;
  Best Local Similarity   99.2%;  Pred. No. 1.3e-148;
  Matches  604;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;
```

```
Qy      58 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 117
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy     118 ACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 177
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy     178 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATG 237
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy     238 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 297
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy     298 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 357
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy     358 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 417
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy     418 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 477
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy     478 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 537
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy     538 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 597
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy     598 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGG 657
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db     539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy     658 GAAAACTAT 666
           |||||||||
Db     599 GAAAACTAT 607
```

RESULT 3
AV338075

```
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .649
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="6430407H24"
                      /clone_lib="RIKEN full-length enriched, adult male
                      olfactory bulb"
                      /sex="male"
                      /tissue_type="olfactory brain"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 100.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                      3']. cDNA was cloned into the XhoI and BamHI sites.
                      Vector: a modified pBluescript KS(+) after bulk excision
                      from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                      BamHI"
BASE COUNT      209 a    105 c    117 g    218 t
ORIGIN
```

```
  Query Match             40.2%;  Score 368;  DB 9;  Length 649;
  Best Local Similarity   98.7%;  Pred. No. 3.6e-90;
  Matches  371;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     540 ATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGA 599
```

```
Db      2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61
Qy    600 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 659
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121
Qy    660 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 719
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181
Qy    720 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 779
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241
Qy    780 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 839
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301
Qy    840 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 899
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361
Qy    900 ATGGTTTTGGAATTAA 915
Db    362 ATGGTTTTGGAATTAA 377

RESULT 4
AV725071
LOCUS       AV725071                  716 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION   AV725071
VERSION     AV725071.1  GI:10830091
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES    Location/Qualifiers
     source 1. .716
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HTCBLB02"
            /clone_lib="HTC"
            /tissue_type="Hypothalamus"
            /dev_stage="Adult"
            /lab_host="SOLR"
            /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT  219 a    113 c    133 g    249 t      2 others
ORIGIN

  Query Match          40.0%;  Score 366;  DB 9;  Length 716;
  Best Local Similarity  91.7%;  Pred. No. 1.3e-89;
  Matches  387;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    494 AAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAAT 553
Db      1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60
Qy    554 TTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATG 613
Db     61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120
Qy    614 CATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATT 673
Db    121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180
Qy    674 ATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAA 733
Db    181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240
Qy    734 AATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAA 793
Db    241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300
Qy    794 AGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACC 853
Db    301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360
Qy    854 ATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 913
Db    361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420
Qy    914 AA 915
Db    421 AA 422

RESULT 5
CNS059ST/c
LOCUS       CNS059ST                  1101 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            045P22 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL327638
VERSION     AL327638.1  GI:8221227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES    Location/Qualifiers
     source 1. .1101
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon 99883"
```

/clone="045P22"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA045DH11A1~end : T3"
BASE COUNT 287 a 249 c 269 g 278 t 18 others
ORIGIN

Query Match 31.8%; Score 291.2; DB 12; Length 1101;
Best Local Similarity 59.3%; Pred. No. 5e-69;
Matches 544; Conservative 0; Mismatches 369; Indels 5; Gaps 3;

```
Qy      2 CAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGA 61
          |||  |    | ||   |||   || |  || ||| | ||| | ||||| || || || |
Db    951 CAACCAAGAGTGATAAAAACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAA 892

Qy     62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAG 121
          ||||  ||||| |     ||| | ||    |||    |  | || ||  |  | ||  | |
Db    891 CCTTCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGG 832

Qy    122 ACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGG 181
          || |     | |  ||||| ||  |||  ||  |  | ||||| |||||||||  | | |
Db    831 ACAGAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTG 772

Qy    182 ATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATT 241
          | ||| | ||| | ||       |   | ||  || | ||||||||||| ||| |  |
Db    771 ACCTGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACA 712

Qy    242 TAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTC 301
          | ||||| ||| | ||| || |      |   ||| || ||   | ||||||||||||| |
Db    711 TGGAGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGC 652

Qy    302 TAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAA 361
          | |  ||||| |  ||||| || ||        ||  |   || |    ||   |||
Db    651 TGAACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTG 592

Qy    362 ATCCCTTTGTG---TTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 418
          || |    | |   ||||     || |||||  | ||  |||| || ||| |||| || | |
Db    591 ATGCTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCA 532

Qy    419 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 478
          |||||||||  || |||||  | || || |  ||||||||  |   ||||  | ||||
Db    531 ACTGGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGG 472

Qy    479 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 538
          |||||||  ||||| | |||||||||  || || |    ||| || |      | || ||||
Db    471 TCCACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCA 412

Qy    539 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 598
          |  | |  || || || || || |||||||| ||||| || |||||||| ||| | || |
Db    411 TCGCCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTG 352

Qy    599 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 658
          |||| || ||||  || ||    |  ||  |||| || || || ||||| ||| || |||
Db    351 AAAAACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGG 292

Qy    659 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 718
          | ||||| |||||    || |  |  ||| | || || || |||||||| |||  |||
Db    291 AGAACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAG 232

Qy    719 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 778
          |||  ||| | ||  | ||| |   |    | ||||||||  |  |  |||||||||  |
Db    231 CCAAGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAAT 172

Qy    779 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 838
          | ||| |   |  |||     ||      ||    |  ||  || |           | |
Db    171 ATTTTCANGTGGCGAAGACACTTTAAAGCAA-GGGCAGTGGCATTTGCCCGAGGCCACCT 113

Qy    839 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAG 897
          | | ||| || ||  | || |  |||| | |  |||| | ||| |   |  | ||| |
Db    112 GTCAGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANAC 53

Qy    898 AAATGGTTTTGGAATTAA 915
          || ||||  |||  || |
Db     52 AAGTGGTACTGGGGTTGA 35
```

RESULT 6
BM021081/c
LOCUS BM021081 579 bp mRNA linear EST 30-OCT-2001
DEFINITION ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;, mRNA sequence.
ACCESSION BM021081
VERSION BM021081.1 GI:16535437
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 579)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T. , Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
High quality sequence stop: 436.
FEATURES Location/Qualifiers
source 1. .579
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."
BASE COUNT 203 a 103 c 93 g 180 t
ORIGIN

Query Match 29.7%; Score 272.2; DB 10; Length 579;
Best Local Similarity 92.6%; Pred. No. 6.8e-64;
Matches 286; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
Qy    607 TACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTAT 666
          |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy    667 GAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAG 726
          |||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||||
```

```
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy    727 TTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAAC 786
          ||||||| |||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy    787 TGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCA 846
          ||||| |||||||| |||||||| || ||||  ||||||||||||||||||||  ||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy    847 TGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 906
          ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy    907 TGGAATTAA 915
          |||||||||
Db    279 TGGAATTAA 271


RESULT 7
CNS0579Z/c
LOCUS       CNS0579Z                1100 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            007H05 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL324368
VERSION     AL324368.1  GI:9557250
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007H05"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA007CD03C1-end : T7"
BASE COUNT      314 a    228 c    252 g    304 t      2 others
ORIGIN

  Query Match             19.6%;  Score 179.6;  DB 12;  Length 1100;
  Best Local Similarity   55.6%;  Pred. No. 2.1e-38;
  Matches  430;  Conservative    0;  Mismatches  331;  Indels   13;  Gaps    4;

Qy     30 CATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAAT 89
          ||| ||| | ||| | ||||| | ||| |||| |||| |||        || ||  |||
Db    774 CATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTTGCATACGACTCTTGTAAAGTTTA 715

Qy     90 GTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGC 149
          | |||||| | || || | |||||| || || ||  | ||| ||||| | || ||
Db    714 CTACAATATTGACGGCTGTGAGCTCACAGATGATCGATCCCTTTACGACAAAGCGCAGGC 655

Qy    150 GGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAG 209
          || ||| |   ||| |  |||||    ||| |||||   ||||| ||   | || |  |
Db    654 TGTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGGAAACTTGCCAGTGGAGCCGCG 595

Qy    210 GCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAA 269
          ||||  |||||||  |||||||||| |  |  | || || ||||   ||||   |   |
Db    594 GCCATATTTTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGAT 535

Qy    270 GAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCA 329
          |||  |||    | |||||| ||||||| |  ||  |    ||||| || ||
Db    534 ACCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGT 475

Qy    330 AGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 386
          ||     ||    |     | | |  | |    |  |     |  | |   ||||
Db    474 AGCCAGGTACCCTCTTACCATCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGA 415

Qy    387 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAG------CATGCCAG 440
          ||  | ||  | |  || ||| | || || ||    |   |              | | ||
Db    414 GAAAAACAAACTTGCCTGTTGGATCGTTAGCAATGCTACATCCACTGGGACAGGTACAAG 355

Qy    441 GGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATT 500
          |   | | ||| || || ||   |||   |||| | ||        | |||    |  |
Db    354 AGCGCAATTTTATAATGAACTATCCAAACATATTAACATTAGTGTTTTTGGTGTGGTGTA 295

Qy    501 CGGAGAATACGTGAACGATAAAAATCTGATT---CCCACCATATCTACTTGTAAATTTTA 557
          |  ||    |  |   || |   ||| |  |   || ||| |  |||  |||| |||||
Db    294 CACAGGGGTCAGGTTGGAGATTGATCAGTATTACCCGACCGTTGCTAGCTGTAAGTTTTA 235

Qy    558 TCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATT 617
          | |||||||| |||||| | ||||||||||||||||||| ||| || ||    |  |
Db    234 CTTATCATTTGAGAACTCATTACACAAAGATTACATCACAGAGAAGGTCAACGNACCCCT 175

Qy    618 TTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATAT 677
          |   || ||  | || ||||| || |||||  ||  | || |  |||| || |  |   |
Db    174 TGCAGCAGGAACCGTTCCTGTAGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTT 115

Qy    678 TCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATA 737
          ||  | || || || ||||| || || ||||||     ||| |  |   | ||| |  |
Db    114 CCCTTCAGACTCCTTTATTCACGTAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCA 55

Qy    738 TCTGAAGGAAGTTGACAA-AAACAATAAGTTGTACCTTAGTTACTTTAACTGGA 790
          ||||       | |||||   |  |  ||   ||| |  | |||||| | ||||
Db     54 TCTGCTCTCCTTGGACAAGGNATGAGGAGGCATACATGCGCTACTTTGAGTGGA 1


RESULT 8
CNS01V8V
LOCUS       CNS01V8V                844 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            198K09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL168808
VERSION     AL168808.1  GI:7806865
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="198K09"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG198AF05SP1~end :
                     PUC-Ori"
BASE COUNT      199 a    190 c    205 g    245 t      5 others
ORIGIN

  Query Match           19.2%;  Score 175.6;  DB 12;  Length 844;
  Best Local Similarity   57.3%;  Pred. No. 2.5e-37;
  Matches  332;  Conservative    2;  Mismatches  242;  Indels    3;  Gaps    1;

Qy      14 ATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTA 73
           ||   |||   || ||  | ||| | ||| | ||||||||||||   ||||||||||  |
Db     269 ATAAAAACCTGACCACGGTGCTGCTCTGGTTCTGGCCATTTGGGATTACCTTTGACCCGA 328

Qy      74 CATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGT 133
             |||   ||    |   ||| | || ||     | ||| | ||| |    |  |
Db     329 ACGTGTGCAGCGCTTACTATGGTATAGACGGCTGTTTCATTACAGCGGACAGGAACTATT 388

Qy     134 ACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACT 193
            || | ||        | || | ||||| ||| ||||:||      || ||| | ||
Db     389 TCAGCGAAGGAATGCGGNTCGTTATTCATCATCGAGAMATTTCTCCCGACCTGTCCAATC 448

Qy     194 TACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCA 253
           | || |||||||   ||||    |  ||:|  ||| | ||| | ||  | |||   ||
Db     449 TGCCGCAGCAGGAGCGGCCGTTGTACCARATGTGGGTCTGGTTCAACATGGAGAATCCTT 508

Qy     254 CTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCC 313
           |  || ||  |         ||| || ||   ||||||||||||||||| | ||||||
Db     509 CGTACTCCGACAGGTTTCTTGGAATGGACGGCTTGTTCAACCTGACTCTGAATTATCGTC 568

Qy     314 GTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGT 373
             ||| | |||||     ||||| |||  | | ||        | || |     |  |
Db     569 TGGATGCTGATATTTCCATGCCTCATGTGTACCTGGTTCCTGAAAAAAGTGAGGAAGACT 628

Qy     374 TTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC 433
           |||     || ||||| || |||||||| |||||| | || || ||||||||     |
Db     629 TTGTCCCCCCCAGCAAAGATAAGTTGGTCTGCTGGATAGTCAGCAACTGGAAGGAGCATT 688

Qy     434 ATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCC 493
           ||    |||| || || | ||| |||||    ||||    ||  | |||  ||| ||||| |
Db     689 ATATACGGGTGAAATACTTCAAGGAGCTGTCCAAGCACATAAAGATCACCACTTATGGAC 748

Qy     494 AAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAAT 553
             | || || |  |||   |||     || |      ||| |||  | |  |||||  |
Db     749 GTGGCTTTGGGGGCTAC---AACTGGGGAAGTGCCTCGTCCATCATGGCCAGCTGTAAGT 805

Qy     554 TTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACA 592
           | || || || ||||| ||||| ||  |||  || ||||
Db     806 TCTACCTGTCTTTTGAGAACTCCATCTACATTGACTACA 844


RESULT 9
```

```
BM082484/c
LOCUS       BM082484                570 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223
            3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;,
            mRNA sequence.
ACCESSION   BM082484
VERSION     BM082484.1  GI:16929414
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN

  Query Match           18.7%;  Score 171.4;  DB 10;  Length 570;
  Best Local Similarity   57.8%;  Pred. No. 3.2e-36;
  Matches  329;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    1;

Qy     217 TTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGC 276
           || ||||| |||||||||| |  |||||||  || || ||  ||||      ||    |||
Db     570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy     277 ATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCT 336
           | ||| || |||||||||  |||||  |  ||  | |  ||| | ||||| | ||| |
Db     510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy     337 TATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAG 396
             | ||    ||  ||  | |    ||  ||  ||||   | || |  |||| ||
Db     450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy     397 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGGGTCAAG 447
           | || || ||| ||||||||||     |||          ||     || |||  || |
Db     390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331
```

```
Qy      448 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 507
Db      330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271
Qy      508 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 567
Db      270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211
Qy      568 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 627
Db      210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151
Qy      628 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 687
Db      150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91
Qy      688 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 747
Db      90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31
Qy      748 GTTGACAAAAACAATAAGTTGTACCTTAG 776
Db      30 CTAAATAAAAATGACATTGCATATCACAG 2

RESULT 10
BB595164
LOCUS       BB595164                 648 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB595164 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION   BB595164
VERSION     BB595164.2  GI:16450167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11491766.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            Watahiki,M., Yoneda,Y., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES    Location/Qualifiers
     source 1. .648
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="C030004N01"
            /clone_lib="RIKEN full-length enriched, adult male corpus
            striatum"
            /sex="male"
            /tissue_type="corpus striatum"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Site_1: SalI; Site_2: BamHI; cDNA library was
            prepared and sequenced in Mouse Genome Encyclopedia
            Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in
            RIKEN. Division of Experimental Animal Research in Riken
            contributed to prepare mouse tissues. 1st strand cDNA was
            primed with a primer [5'
            GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse
            transcriptase and subsequently enriched for full-length by
            cap-trapper. cDNA went through one round of normalization
            to Rot = 10.0 and subtraction to Rot = 185.0. Second
            strand cDNA was prepared with the primer adapter of
            sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
            3']. cDNA was cloned into the XhoI and BamHI sites.
            Vector: a modified pBluescript KS(+) after bulk excision
            from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
            BamHI"
BASE COUNT  141 a    206 c    138 g    163 t
ORIGIN

  Query Match          18.7%;  Score 170.8;  DB 9;  Length 648;
  Best Local Similarity  98.9%;  Pred. No. 4.8e-36;
  Matches  172;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
Db      475 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 534
Qy       61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
Db      535 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAACA 594
Qy      121 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATC 174
Db      595 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATATAGACATC 648

RESULT 11
CNS02Q0S/c
LOCUS       CNS02Q0S                 887 bp    DNA     linear   GSS 14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            156L20 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL208693
VERSION     AL208693.1  GI:7867512
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
```

```
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 887)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .887
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="156L20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG156DF10LP1~end : T7"
BASE COUNT      232 a    188 c    206 g    253 t      8 others
ORIGIN

  Query Match             18.0%;  Score 164.8;  DB 12;  Length 887;
  Best Local Similarity   54.8%;  Pred. No. 2.3e-34;
  Matches  385;  Conservative    5;  Mismatches  297;  Indels   15;  Gaps    3;

Qy      30 CATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAAT 89
           ||| ||| | ||| | ||||| | ||| |||| || |         || ||  |||
Db     704 CATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTGGCATACGACTCTTGTAAAGTTGA 645

Qy      90 GTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGC 149
           : |||||| |  | ||  | ||:|||   || || ||  |  || ||||| | || ||
Db     644 CKACAATATTGACAGCTGTGAGCTSACAGATGATCGATCCCTTGACGACAAAGCGCAGGC 585

Qy     150 GGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAG 209
           || ||| |   ||| |  |||||    ||| ||||| | ||||| ||   | || |  |
Db     584 TGTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGCAAACTTGCCAGGGGAGCCGCG 525

Qy     210 GCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAA 269
           ||||   |:||||  |||||||| |  |  | || || ||||   ||||   |   |
Db     524 GCCATATTKTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGAT 465

Qy     270 GAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCA 329
           ||| | |||    | |||||| ||||||| |  ||  |    ||||| || ||
Db     464 ACCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGT 405

Qy     330 AGTGCCTTA---TGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 386
           |     ||   |     | | |  | |   | |    |         | |   ||||
Db     404 GGCCAGGTACCCTCTTACCATCAGAGAGGAAAGACTGACGGAGAAAATAGKTCTGCCGGA 345

Qy     387 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCA- 445
           ||  | ||  | |||||||| | || || ||          |:  || || || ||
Db     344 GAAAAACAAAATTGTGTGCTGGATCGTCAGCAAAACTGCAAAAAASACTGGCACGGGCAC 285

Qy     446 --------AGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGC 497
                   | |  |     || ||   |||| |||| |  |  |    | |||    ||
Db     284 AAGGGAACAATTCTTTGGTGAACTATCCAAGCATATTCATGTTGATGTTTTTGGTGTGGC 225

Qy     498 ATTCGGAGAATACGTGAACGATAAAAATCTGATT---CCCACCATATCTACTTGTAAATT 554
           |  || |     |   || |  ||| | |  || ||||| |||  ||||| ||
Db     224 GGACACAGGAGTGAAGTTGGAGATTGATCAGTATTACCCGACCATTGCTAGCTGTAAGTT 165

Qy     555 TTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGC 614
           ||| || || ||||| ||||||   |||||||||:|||||||||| ||| || ||    |
Db     164 TTACCTGTCCTTTGAGAACTCATAACACAAAGATKACATCACAGAGAAGGTCAACGGACC 105

Qy     615 ATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTA 674
           ||   |  ||  | || ||||| || ||||| ||  | || .|  ||||| || |  |
Db     104 CCTTGCAGTAGGAACCGTTCCTGTGGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTT 45

Qy     675 TATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTC 716
           | ||  | || || || ||||| || || ||||||   ||
Db      44 CTTCCCTTCAGACTCCTTAATTCACGTAGATGATTTTCTTTC 3


RESULT 12
CNS0429P/c
LOCUS       CNS0429P                1016 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            075P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271222
VERSION     AL271222.1  GI:7993186
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1016)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1016
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="075P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG075DH01SP1~end :
                     PUC-Ori"
BASE COUNT      246 a    216 c    253 g    298 t      3 others
ORIGIN

  Query Match             17.1%;  Score 156.8;  DB 12;  Length 1016;
  Best Local Similarity   55.8%;  Pred. No. 3.8e-32;
  Matches  386;  Conservative    0;  Mismatches  292;  Indels   14;  Gaps    4;

Qy      32 TTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGT 91
           ||||| | ||| | ||||| | ||   |  ||||| | |     |||    |    |
Db     715 TTCTGCTGTGGTTCTGGCCCATCGGCACGGTGTTTGATTTCAAGGATTGCTCCACCCACT 656
```

```
Qy      92 TCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGG 151
           |||| ||  |  | ||  ||||| ||     ||| | ||  |||||| |||| |  | |  |
Db     655 TCAAGATTGACAGCTG-CATCTGACGGACGACAGATCTCTGTACAGCAAAGCAAAGGGAG 597

Qy     152 TCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGC 211
           ||  |  |  |||| | |||||  |  ||| |      | ||| ||   |      |  | | |
Db     596 TCATTTTCTACCACAAAGACACAACATGGCACACAGACAGCTTGCCCACGGGTCCACGCC 537

Qy     212 CACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGA 271
           |  |  || ||||  ||| |||||  | ||| | || || || ||  ||||   |   ||
Db     536 CCCATTTCCAGAGGTGGGTTTGGTTTAATGTGGAATCCCCAACAAACACATACAGGAAAC 477

Qy     272 GTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAG 331
           ||  | ||| || ||||||||||||| || |  ||  |  | ||   |  ||     |
Db     476 CAGGTCTGGAAAACCTGTTCAACCTGACGCTGAGCTACAGGAGAGACGGCGGCATTGTGG 417

Qy     332 TGCCTTATGGCTTCTTGACGGTGAGCACA---AATCCCTTTGTGTTTGAAGTGCCAAGCA 388
           |    | |   | | |          | ||   |  ||   | ||||    |||| |  |
Db     416 CGAGGAACGAGATGTCTATCCGAGACCCAGGTGACACCGAGGACTTTGTTCTGCCCAAGA 357

Qy     389 AGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATG---------CCA 439
           | || |||  |||||| ||| | || || ||| |||| |         |||
Db     356 AAGACAAGACGGTGTGTTGGATCGTCAGCAACAATGACCCGAGAACAGGAACAGGCACCA 297

Qy     440 GGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCAT 499
           | | | | || |||   || ||    |||    ||  |  |  ||   |  ||| |||| |
Db     296 GAGACCAATACTACCGTGACCTGTCCAAACACATCAAGGTGGACGTTTGGGGCAAAGCCT 237

Qy     500 TCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATC 559
           | ||  | | | |||     |  |   |       ||| || ||  | |  |||||||| ||||
Db     236 TTGGGAATTTCCTGAGTTTTGAGGAATACTATCCAACAATTGCCAGCTGTAAATTCTATC 177

Qy     560 TTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTT 619
           | || || || ||||| || ||||||||  ||||| || || ||| || ||     | |||
Db     176 TGTCTTTCGAGAACTCCATCCACAAAGACTACATGACGGAGAAGATCAACGGGCCCTTTG 117

Qy     620 TGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATAT-T 678
           |||||||  |||||  ||| | ||  |||| ||  | |   ||||  |  | ||   | | |
Db     116 TGGCTGGGTCAGTTCCTATAGTTTTGGGACCTCCAAAAAAAAAAAAAAAAAAAAATTCTAT 57

Qy     679 CCAGCTGATTCATTCATTCATGTGGAAGATTT 710
           ||  | |||||  || ||||| ||  | || ||
Db      56 CCTTCCGATTCGTTTATTCACGTTAACGACTT 25
```

RESULT 13
BB398103
LOCUS BB398103 199 bp mRNA linear EST 15-JUL-2000
DEFINITION BB398103 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330006N11 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB398103
VERSION BB398103.1 GI:9217499
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 199)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .199
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C330006N11"
/clone_lib="RIKEN full-length enriched, ES cells"
/cell_type="ES cells"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT 56 a 37 c 44 g 62 t
ORIGIN

Query Match 16.6%; Score 152.2; DB 9; Length 199;
Best Local Similarity 92.5%; Pred. No. 4.5e-31;
Matches 160; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy     743 AGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTA 802
           |||  || |||    ||||| |||| ||| ||||||||||||||||||| |||||||||
Db       6 AGGCTGTGGACCCCTACAATCAGTTCTACTTTAGTTACTTTAACTGGAGAGAGGATTTTA  65

Qy     803 CTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAA 862
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      66 CTGTAAACCTACCACGGTTTTGGGATTCACATGCATGCCTGGCATGCGACCATGTAAAAA 125

Qy     863 GGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 GGCATCTAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 178
```

RESULT 14
BB649967
LOCUS BB649967 625 bp mRNA linear EST 26-OCT-2001
DEFINITION BB649967 RIKEN full-length enriched, 16 days embryo head Mus musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION BB649967

VERSION BB649967.1 GI:16484222
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 625)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .625
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C130099N17"
/clone_lib="RIKEN full-length enriched, 16 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"
BASE COUNT 136 a 198 c 135 g 156 t
ORIGIN

Query Match 16.4%; Score 150; DB 9; Length 625;
Best Local Similarity 100.0%; Pred. No. 2.4e-30;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 535

Qy     61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    536 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 595

Qy    121 GACCGCTCATTGTACAACAAATCCCATGCG 150
          |||||||||||||||||||||||||||||
Db    596 GACCGCTCATTGTACAACAAATCCCATGCG 625
```

RESULT 15
BJ050300/c
LOCUS BJ050300 666 bp mRNA linear EST 06-DEC-2001
DEFINITION BJ050300 NIBB Mochii normalized Xenopus neurula library Xenopus laevis cDNA clone XL023o15 3', mRNA sequence.
ACCESSION BJ050300
VERSION BJ050300.1 GI:17392988
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 666)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .666
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL023o15"
/clone_lib="NIBB Mochii normalized Xenopus neurula library"
/tissue_type="whole embryo"
/dev_stage="stage 15"
BASE COUNT 178 a 116 c 157 g 215 t
ORIGIN

Query Match 16.2%; Score 148.4; DB 10; Length 666;
Best Local Similarity 55.3%; Pred. No. 6.8e-30;
Matches 354; Conservative 0; Mismatches 276; Indels 10; Gaps 3;

```
Qy    251 CCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATC 310
          | |  ||||  ||| | |         || ||  |  |  |||| |||||  |  |||| |
Db    666 CAAGCCACAATCCCAACACAGCGTTTATGGAGAAACTCATCAATCTGACCATGTCTTACC 607

Qy    311 GCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTG 370
          |   ||| ||||||||       ||||||||||   |      |
Db    606 GTGCTGACTCAGATATTTTTACTCCTTATGGCTGGCTAGAAAGGCATGATGGAAAAGAAA 547
```

```
Qy    371 TGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTG 430
Db    546 ACTTCACTATGCCCGCCAAGACAAAACTAGTAGCCTGGGTAGTCAGTAACTGGAACCCAA 487

Qy    431 AGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATG 490
Db    486 ACTCCAGAAGAATCAAATATTATGAGGAGTTAAAGCCACACCTTGCTGTTGATATCTATG 427

Qy    491 GCC--AAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTG 548
Db    426 GCCTACAGCATTTACCGCTTCCTAGACC-----AAAACACATTGAAATTCTTTCCCAGTA 372

Qy    549 TAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT- 607
Db    371 CAAATTTTATTTAGCTTTTGAAAATTCAATTCACGAAGACTACATCACTGAGAAACTCTG 312

Qy    608 --ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTA 665
Db    311 GCATAATGCCCTCTCATCGTGGGCTCTACCTGTTGTTTTGGGGCCCTCGCGTTCTAACTA 252

Qy    666 TGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGA 725
Db    251 TGAGCGTTTTTTACCCCCCGAGTCATTCATCCACGTAGATGACTTTCCAACTGCTAAGGG 192

Qy    726 GTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAA 785
Db    191 ACTTGCTGATTATCTAAACGAGCTTGATAAAGATGATAATAAATACATGCAACATTTCAA 132

Qy    786 CTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGC 845
Db    131 CTGGAGGAAACATCTAAAACCAGTAGCACCTGAGAGTTGGAACCCCCATTACTGCAAAGT 72

Qy    846 ATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTT 885
Db     71 ATGCAAAGCTATAAAAGACGCACCGTCTTATAGGACTTTT 32
```

Search completed: October 6, 2002, 07:15:23
Job time : 774.105 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 18.6694 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-3_COPY_280_1194
Perfect score: 915
Sequence: 1 acaaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128.6	14.1	1814	2	US-08-483-151-1	Sequence 1, Appli
2	128.6	14.1	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	128.2	14.0	1086	1	US-07-914-281-12	Sequence 12, Appl
4	128.2	14.0	1086	1	US-08-393-246-12	Sequence 12, Appl
5	128.2	14.0	1086	1	US-08-525-058A-12	Sequence 12, Appl
6	128.2	14.0	1086	2	US-08-696-731-12	Sequence 12, Appl
7	128.2	14.0	1086	4	US-09-042-531-12	Sequence 12, Appl
8	128.2	14.0	2042	3	US-09-063-237-2	Sequence 2, Appli
9	128.2	14.0	2043	1	US-07-914-281-1	Sequence 1, Appli
10	128.2	14.0	2043	1	US-08-393-246-1	Sequence 1, Appli
11	128.2	14.0	2043	1	US-08-273-411-4	Sequence 4, Appli
12	128.2	14.0	2043	1	US-08-525-058A-1	Sequence 1, Appli
13	128.2	14.0	2043	2	US-08-696-731-1	Sequence 1, Appli
14	128.2	14.0	2043	4	US-09-042-531-1	Sequence 1, Appli
15	128.2	14.0	2043	5	PCT-US91-00899-1	Sequence 1, Appli
16	126.6	13.8	1316	1	US-07-914-281-10	Sequence 10, Appl
17	126.6	13.8	1316	1	US-08-393-246-10	Sequence 10, Appl
18	126.6	13.8	1316	1	US-08-525-058A-10	Sequence 10, Appl
19	126.6	13.8	1316	2	US-08-696-731-10	Sequence 10, Appl
20	126.6	13.8	1316	4	US-09-042-531-10	Sequence 10, Appl
21	124.6	13.6	1654	1	US-07-914-281-13	Sequence 13, Appl
22	124.6	13.6	1654	1	US-08-393-246-13	Sequence 13, Appl
23	124.6	13.6	1654	1	US-08-525-058A-13	Sequence 13, Appl
24	124.6	13.6	1654	2	US-08-696-731-13	Sequence 13, Appl
25	124.6	13.6	1654	4	US-09-042-531-13	Sequence 13, Appl
26	120	13.1	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	120	13.1	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	120	13.1	1488	1	US-07-914-281-9	Sequence 9, Appli
29	120	13.1	1488	1	US-08-393-246-9	Sequence 9, Appli
30	120	13.1	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	120	13.1	1488	2	US-08-696-731-9	Sequence 9, Appli
32	120	13.1	1488	4	US-09-042-531-9	Sequence 9, Appli
33	120	13.1	2175	4	US-08-482-073-9	Sequence 9, Appli
34	120	13.1	2861	4	US-08-482-073-10	Sequence 10, Appl
35	120	13.1	3647	1	US-07-914-281-7	Sequence 7, Appli
36	120	13.1	3647	1	US-08-393-246-7	Sequence 7, Appli
37	120	13.1	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	120	13.1	3647	2	US-08-696-731-7	Sequence 7, Appli
39	120	13.1	3647	4	US-09-042-531-7	Sequence 7, Appli
40	120	13.1	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	116.8	12.8	2134	2	US-08-483-151-3	Sequence 3, Appli
42	116.8	12.8	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	33.8	3.7	1730	1	US-07-817-920-1	Sequence 1, Appli
c 44	33.8	3.7	1730	1	US-08-117-006-1	Sequence 1, Appli
c 45	33.8	3.7	1730	1	US-08-216-594-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 14.1%; Score 128.6; DB 2; Length 1814;
Best Local Similarity 52.5%; Pred. No. 5.9e-30;
Matches 331; Conservative 0; Mismatches 294; Indels 6; Gaps 2;

```
Qy    104 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 163
          | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy    164 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 223
          |  | ||  |          |   ||  | |||| | | |    |||||||  |   |||
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    224 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 283
            ||| | |||       | || || |||| | | ||||         | || |
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy    284 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
          | | ||||||  |   || |  ||||| ||||||||||||||   || || || ||
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy    344 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 403
          ||||  |     | |   |||      |   | | ||   |||    | | |||
Db    788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy    404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 463
          ||||||  | || || |     |      |||     | |  |||   |||    ||||
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy    464 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 523
          |      | |  |  |  |     |  || |  ||   ||||      | |   || |||
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy    524 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 583
          ||||  |||||  |   |     |      || || ||  | ||||| ||||||   ||
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    584 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 640
            || |||||||| || ||| |||     ||||||  |   ||||||| | ||||| || |
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    641 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 700
          |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    701 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 731
          |||| || || | |||| || ||||  | ||
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;    APPLICANT:  The General Hospital Corporation
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             14.1%;  Score 128.6;  DB 5;  Length 1814;
  Best Local Similarity   52.5%;  Pred. No. 5.9e-30;
  Matches  331;  Conservative    0;  Mismatches  294;  Indels    6;  Gaps    2;

Qy    104 GGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACC 163
          | |||| ||| |   |  ||||     ||     ||   |  |||| ||  |  | ||||
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy    164 ATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGA 223
          |  | ||  |          |   ||  | |||| | | |    |||||||  |   |||
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    224 AATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAAC 283
            ||| | |||       | || || |||| | | ||||         | || |
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy    284 ACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
          | | ||||||  |   || |  ||||| ||||||||||||||   || || || ||
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy    344 TCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGT 403
          ||||  |     | |   |||      |   | | ||   |||    | | |||
Db    788 GCTTGGAGCCTCTCTCTGGGCCC---ACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy    404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCA 463
          ||||||  | || || |     |      |||     | |  |||   |||    ||||
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy    464 GCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAA 523
          |      | |  |  |  |     |  || |  ||   ||||      | |   || |||
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy    524 ATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACA 583
          ||||  |||||  |   |     |      || || ||  | ||||| ||||||   ||
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    584 AAGATTACATCACAGAAAAGCTCTA---CAATGCATTTTTGGCTGGTTCAGTACCTGTTG 640
            || |||||||| || ||| |||     ||||||  |   ||||||| | ||||| || |
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    641 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 700
          |||||  ||  || |||  | ||| |||  || | | ||| | ||| | ||  | || |
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    701 TGGAAGATTTTAACTCTCCCAGTGAGTTAGC 731
          |||| || || | |||| || ||||  | ||
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 3
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;       ADDRESSEE:  P.C.
;       STREET:  1755 Jefferson Davis Highway, Fourth Floor
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/914,281
;       FILING DATE:  19920720
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1086 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             14.0%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 6e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          | || ||||  |   |   |  | | |   | | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
             ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
             |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || |||||  |||||  |||   |||||||| || ||    ||||||||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||          | |  |  |  ||   || || |   |   |     |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
          | |    |  ||-||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || ||||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 4
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             14.0%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 6e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          | || ||||  |   |   |  | | |   | | || |       | |||||  |     |||
```

```
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282

Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342

Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453

Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513

Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573

Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630

Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 5
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             14.0%;  Score 128.2;  DB 1;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 6e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165

Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222

Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282

Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342

Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453

Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513

Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573

Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630

Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 6
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
```

```
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             14.0%;  Score 128.2;  DB 2;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 6e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||     |  |  |  |||||  |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          | || ||||  |   |   |  | | |   | | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
            ||||| ||| | || || ||| ||||| || ||  ||  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | | |    |   ||           | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || |||||  |||||  |||   |||||||| || ||    ||||||||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
            | |    |   ||||    ||  | ||    |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || |||||||| || |||||      || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |   |||| ||  | ||    || ||| |||| |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 7
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match             14.0%;  Score 128.2;  DB 4;  Length 1086;
  Best Local Similarity   53.0%;  Pred. No. 6e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          ||||| |||| | ||||||    |||||     |  |  |  ||||| |  | |||||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          | || ||||  |   |   |  | |   | | || |       | |||||  |      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
          |||||  ||| | ||  ||  |||  ||||| || ||  ||  ||   |    | | |  ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  |||||  ||  ||  |  | ||  |||  | ||  ||  ||  |||      |||  ||  |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |     |    ||         | |  |  |  |   ||||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||         | |  |  |  ||   || || |   |   |     |   | ||
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 687

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
          | |    |   ||||    ||  | ||     |  || || || ||  | || || |||
Db    688 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |   |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    |||||  ||||  |   | |||  | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    ||||  ||  | ||    || |||  ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 8
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             14.0%;  Score 128.2;  DB 3;  Length 2042;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||     |  |  |  |||||  |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          | || ||||  |   |   |  | |   | | || |       | |||||  |      |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
          |||||  ||| | ||  ||  |||  ||||| || ||  ||  ||   |    | | |  ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  |||||  ||  ||  |  | ||  |||  | ||  ||  ||  |||      |||  ||  |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |     |    ||         | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
          || |||||  |||||  |||   |||||||| || ||    ||||||||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||         | |  |  |  ||   || || |   |   |     |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
          | |    |   ||||    ||  | ||     |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |   |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    |||||  ||||  |   | |||  | ||  |
```

```
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 9
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             14.0%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
          | || ||||  |   |   |  | |   | | || |      | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
          |||||  ||| | || ||  ||| |||||  || ||  ||  ||   |  | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
          || |||||  |||||  |||   |||||||| || ||    ||||||||     || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
          |    ||          | |  |  |  ||   || || |   |   |   |    | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
          | |    |  ||||    ||  | ||    |  || || || ||  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
          ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 10
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
```

```
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             14.0%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||    |  |  |  |||||  |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
            || |||| |  |   |  | | |  | | || |     | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
              ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |     |    ||          | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
            || |||||  |||||  |||   |||||||| || ||    |||||||||     || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
            |    ||          | |  |  |  ||   || || |   |   |    |   | ||
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
            | |    |  ||||    ||  | ||     |  || || || ||  |  | || || |||
Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
            ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
            || || || || ||||| ||    ||    ||||| |||  |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
            ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 11
US-08-273-411-4
; Sequence 4, Application US/08273411
; Patent No. 5625124
;  GENERAL INFORMATION:
;    APPLICANT:  Falk, Per
;    APPLICANT:  Gordon, Jeffrey I.
;    TITLE OF INVENTION:  Animal Model for Gastro-Intestinal
;    TITLE OF INVENTION:  Disease
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/273,411
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 815-6508
;      TELEFAX:  (404) 815-6555
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..361
;      OTHER INFORMATION:  /note= "Nucleotides 73 through 1158 encode the GDP-L-fuco
;    PUBLICATION INFORMATION:
;      AUTHORS:  Kukowska-Latallo, et al.
;      JOURNAL:  Genes & Development
;      VOLUME:  4
;      PAGES:  1288-1303
;      DATE:  1990
;      RELEVANT RESIDUES IN SEQ ID NO:  4: FROM 1 TO 2043
US-08-273-411-4

  Query Match             14.0%;  Score 128.2;  DB 1;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
          |||||  ||||  | ||||||    |||||    |  |  |  |||||  |  | |||||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
            || |||| |  |   |  | | |  | | || |     | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
              ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
          |   ||  |  | |  |     |    ||          | |  |  |  |   ||||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642
```

```
Qy     394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || ||||| |||||  |||   |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
           |    ||          | |  | |  ||  || || |   |    |    |    | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
           | |    |  ||||    ||  | ||    |  || || || ||  |  || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
           ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
           || || || |||| ||    ||    ||||| |||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||| |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 12
US-08-525-058A-1
; Sequence 1, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,058A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO

US-08-525-058A-1

Query Match 14.0%; Score 128.2; DB 1; Length 2043;
Best Local Similarity 53.0%; Pred. No. 8.3e-30;
Matches 374; Conservative 0; Mismatches 313; Indels 18; Gaps 4;

```
Qy     106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
           |||||  ||||  | ||||||    |||||    |  |  |  ||||| |  | |||||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
           | || ||||  |   |   |  | |   | | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
           |||||  ||| | || || ||| ||||| || ||  ||  ||   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
           |  ||||| || ||  |  | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
           |   ||  |  | |  |    |    ||          | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || |||||  |||| |||    |||||||| || ||    ||||||||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
           |    ||          | |  | |  ||  || || |   |    |    |    | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
           | |    |  ||||    ||  | ||    |  || || || ||  |  || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
           ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
           || || || |||| ||    ||    ||||| |||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||| |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
           |||||  || |       ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

RESULT 13
US-08-696-731-1
; Sequence 1, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.

```
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/696,731
;       FILING DATE:  14-AUG-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2043 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     ANTI-SENSE:  NO
US-08-696-731-1

  Query Match             14.0%;  Score 128.2;  DB 2;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
           ||||| |||| | ||||||    |||||     |  |  |  ||||| |  | ||||| 
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222
           | || |||| |   |   |  | | |   | | || |      | ||||| |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282
           ||||| ||| | || || ||| ||||| || ||  ||  ||   |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
           |  ||||| || ||   | || ||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393
           |   ||  |  | |  |    |   ||          | |  |  |  |   ||||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453
           || |||||  |||||  |||   |||||||| || ||     ||||||||     || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513
           |  ||          | |  |  |  ||   || || |   |   |    |   | ||
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy     514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573
           | |    |   ||||    ||  | ||     |  || || || ||  | || || |||
Db     760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630
           ||  | |||   || |||||||| || ||||| |     || ||  |  |||  |  |

Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690
           || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750
           ||||| || ||||| || ||  |    ||||  ||  | ||    || ||| |||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
           |||||  || |      ||||| || ||||||  |||| |  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-09-042-531-1
; Sequence 1, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-09-042-531-1

  Query Match             14.0%;  Score 128.2;  DB 4;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165
```

```
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 15
PCT-US91-00899-1
; Sequence 1, Application PC/TUS9100899
;  GENERAL INFORMATION:
;    APPLICANT:  Lowe, John B.
;    TITLE OF INVENTION:  Method and Products For the Synthesis of
;    TITLE OF INVENTION:  Oligosaccharide Structures on Glycoproteins, Glycolipids,
;    TITLE OF INVENTION:  or as Free Molecules, and For the Isolation of Cloned
;    TITLE OF INVENTION:  Genetic Sequences That Determine These Structur
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00899
;      FILING DATE:  19910214
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye Ph.D., Jean-Paul
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-021-55 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-5940
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Blood
;      CELL LINE:  A431
PCT-US91-00899-1

  Query Match             14.0%;  Score 128.2;  DB 5;  Length 2043;
  Best Local Similarity   53.0%;  Pred. No. 8.3e-30;
  Matches  374;  Conservative    0;  Mismatches  313;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCAT 165

Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 AGAGACATCAGCTGGGA---TCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAG 222

Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAA 282

Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342

Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTGAGCACAAATCCCTTTG---------TGTTTGAAGTGCCAAGCAAGGAG 393

Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAGTATTAC 453

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAATACGTG 513

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACA---AGCCCCTG 759

Qy    514 AACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTTGAAAAC 573

Db    760 CCCAAGGGGACCATGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATTCACAAAGATTACATCACAGAAAAGCTCT---ACAATGCATTTTTGGCTGGTTCA 630

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCA 690

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAAGTT 750

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 795
```

```
          |||||  || |       ||||| || ||||||  |||| |  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:20:27
Job time : 29.6694 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 3536.27 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-4
Perfect score: 2822
Sequence: 1 aagctttgctactcagagag.....aacatgttgcgattgaattc 2822

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	2814	99.7	80247	9	AL512406	AL512406 Human DNA
2	1542	54.6	3019	9	AB023021	AB023021 Homo sapi
3	1537.2	54.5	2501	9	HSA238701	AJ238701 Homo sapi
4	1103.2	39.1	1705	10	AF230460	AF230460 Cricetulu
5	1092.2	38.7	2139	10	AB015426	AB015426 Mus muscu
6	1085.4	38.5	2156	10	AB049819	AB049819 Rattus no
7	937.6	33.2	1128	10	AF345993	AF345993 Rattus no
8	823.2	29.2	1707	5	AB035906	AB035906 Gallus ga
9	676.8	24.0	1080	5	AB035905	AB035905 Xenopus l
10	644	22.8	71396	2	AC100674	AC100674 Mus muscu
c 11	557.8	19.8	561	11	G55571	G55571 SHGC-100926
c 12	281	10.0	71396	2	AC100674	AC100674 Mus muscu
13	279.6	9.9	1567	5	AB023627	AB023627 Danio rer
14	234.6	8.3	1832	5	AB023628	AB023628 Danio rer
15	222.6	7.9	1092	5	AB035907	AB035907 Xenopus l
16	220.2	7.8	1264	5	AF288369	AF288369 Gallus ga
17	176.2	6.2	2657	5	GGU73678	U73678 Gallus gall
18	145.2	5.1	1429	10	CGU78737	U78737 Cricetulus
19	145.2	5.1	1462	10	AF090450	AF090450 Cricetulu
20	143.6	5.1	1718	10	AF090449	AF090449 Cricetulu
21	137.2	4.9	1044	5	AB035908	AB035908 Gallus ga
22	132	4.7	2293	4	BTA134FTG	X87810 B.taurus al
23	132	4.7	12607	4	BTA132772	AJ132772 Bos tauru
24	127.6	4.5	1158	10	AB039126	AB039126 Mus muscu
25	127	4.5	1487	3	AF016899	AF016899 Schistoso
26	127	4.5	1814	6	AR028686	AR028686 Sequence
27	127	4.5	3594	10	MMU45980	U45980 Mus musculu
c 28	127	4.5	117993	2	AC091284	AC091284 Mus muscu
29	125.6	4.5	1613	4	BTA132776	AJ132776 Bos tauru
30	125.6	4.5	1619	4	BTA132775	AJ132775 Bos tauru
31	125.6	4.5	1644	4	BTA132774	AJ132774 Bos tauru
32	125.6	4.5	1873	4	BTA132773	AJ132773 Bos tauru
33	124.6	4.4	1119	9	AF345881	AF345881 Macaca mu
34	122.8	4.4	1170	10	AB039127	AB039127 Mus muscu
35	122.8	4.4	1170	10	AB039131	AB039131 Mus muscu
36	122.8	4.4	1615	10	MMU33457	U33457 Mus musculu
37	122	4.3	1113	9	AF345884	AF345884 Saimiri s
38	121.2	4.3	1170	10	AB039124	AB039124 Mus muscu
39	121.2	4.3	1170	10	AB039125	AB039125 Mus muscu
40	121.2	4.3	1170	10	AB039128	AB039128 Mus muscu
41	121.2	4.3	1170	10	AB039129	AB039129 Mus muscu
42	121.2	4.3	1170	10	AB039130	AB039130 Mus muscu
43	121.2	4.3	1170	10	AB039132	AB039132 Mus muscu
44	119.8	4.2	1119	9	AF345883	AF345883 Macaca mu
45	118.4	4.2	1170	10	AB039133	AB039133 Mus spici

ALIGNMENTS

RESULT 1
AL512406
LOCUS AL512406 80247 bp DNA linear PRI 27-APR-2001
DEFINITION Human DNA sequence from clone RP11-504J9 on chromosome 6, complete sequence.
ACCESSION AL512406
VERSION AL512406.14 GI:13897154
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 80247)
AUTHORS Bates,K.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Apr 30, 2001 this sequence version replaced gi:13396709.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
RP11-504J9 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-504J9 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP3-381A2 is at 80148 in this sequence.
The true right end of clone RP11-77B15 is at 100 in this sequence.

FEATURES Location/Qualifiers

source	1. .80247
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="RP11-504J9"
	/clone_lib="RPCI-11.2"
repeat_region	184. .1422
	/note="L2 repeat: matches 875. .2419 of consensus"
repeat_region	1744. .1834
	/note="MIR repeat: matches 166. .258 of consensus"
repeat_region	2950. .2977
	/note="14 copies 2 mer ca 100% conserved"
repeat_region	3411. .3669
	/note="LTR16C repeat: matches 93. .364 of consensus"
repeat_region	4372. .4704
	/note="MLT1A1 repeat: matches 5. .365 of consensus"
misc_feature	complement(7821. .8406)
	/note="match: GSS: Em:AQ550748"
repeat_region	8126. .8184
	/note="MER58B repeat: matches 272. .332 of consensus"
repeat_region	8185. .8696
	/note="MLT2D repeat: matches 1. .513 of consensus"
repeat_region	8779. .8817
	/note="MLT2D repeat: matches 504. .542 of consensus"
repeat_region	8819. .8974
	/note="MER58A repeat: matches 6. .162 of consensus"
misc_feature	complement(9443. .9891)
	/note="match: GSS: Em:B90633"
misc_feature	complement(9452. .9883)
	/note="match: GSS: Em:AQ119695"
repeat_region	9782. .9894
	/note="MIR repeat: matches 92. .205 of consensus"
repeat_region	10089. .10212
	/note="MIR repeat: matches 135. .262 of consensus"
misc_feature	complement(10093. .10671)
	/note="match: GSS: Em:AQ237061"
repeat_region	10223. .10282
	/note="MIR repeat: matches 203. .262 of consensus"
misc_feature	complement(10315. .10631)
	/note="match: GSS: Em:AQ200259"
misc_feature	10805. .11186
	/note="match: GSS: Em:AQ072353"
repeat_region	11657. .11844
	/note="L2 repeat: matches 2546. .2750 of consensus"
misc_feature	complement(11659. .12084)
	/note="match: GSS: Em:AQ518361"
misc_feature	complement(11674. .11987)
	/note="match: GSS: Em:AQ094678"
misc_feature	12083. .12588
	/note="match: GSS: Em:AQ691235"
repeat_region	12596. .12810
	/note="MIR repeat: matches 23. .248 of consensus"
repeat_region	12920. .12986
	/note="L1MA5 repeat: matches 6228. .6295 of consensus"
repeat_region	12987. .13169
	/note="L1MA5 repeat: matches 5852. .6043 of consensus"
repeat_region	13168. .13280
	/note="L1MA8 repeat: matches 5710. .5832 of consensus"
repeat_region	13851. .17218
	/note="L1PB1 repeat: matches 2787. .6155 of consensus"
repeat_region	18903. .18953
	/note="MLT2CB repeat: matches 439. .501 of consensus"
repeat_region	18984. .19433
	/note="MLT2CB repeat: matches 1. .460 of consensus"
misc_feature	complement(19108. .19690)
	/note="match: GSS: Em:AQ563935"
repeat_region	20683. .20801
	/note="AluJo/FRAM repeat: matches 169. .288 of consensus"
repeat_region	20802. .20862
	/note="AluJ/FLAM repeat: matches 3. .66 of consensus"
repeat_region	20912. .21033
	/note="MIR repeat: matches 114. .237 of consensus"
repeat_region	21047. .21128
	/note="41 copies 2 mer ta 63% conserved"
repeat_region	21676. .21859
	/note="L2 repeat: matches 2130. .2308 of consensus"
repeat_region	21890. .23612
	/note="L1P repeat: matches 841. .2570 of consensus"
repeat_region	23601. .26609
	/note="L1PA5 repeat: matches 3107. .6145 of consensus"
repeat_region	26620. .26789
	/note="MER5A repeat: matches 1. .174 of consensus"
repeat_region	28521. .28627
	/note="MER39 repeat: matches 5. .111 of consensus"
repeat_region	28656. .28837
	/note="MER39b repeat: matches 54. .235 of consensus"
repeat_region	28838. .29142
	/note="AluSc repeat: matches 4. .308 of consensus"
repeat_region	29143. .29494
	/note="MER39b repeat: matches 235. .549 of consensus"
repeat_region	30529. .31344
	/note="L1PA3 repeat: matches 5331. .6146 of consensus"
repeat_region	32234. .32340
	/note="L2 repeat: matches 2627. .2750 of consensus"
repeat_region	32399. .32578
	/note="MIR repeat: matches 12. .203 of consensus"
repeat_region	32816. .32979
	/note="MIR repeat: matches 91. .259 of consensus"
misc_feature	complement(33006. .33450)
	/note="match: GSS: Em:AQ261227"
misc_feature	complement(33058. .33444)
	/note="match: GSS: Em:AQ040453"
repeat_region	33312. .33596
	/note="L1M4 repeat: matches 3473. .3773 of consensus"
repeat_region	34856. .34915
	/note="30 copies 2 mer aa 68% conserved"
misc_feature	complement(34987. .35547)
	/note="match: GSS: Em:AQ321598 match: STS: Em:G55571"
misc_feature	complement(36488. .36936)
	/note="match: GSS: Em:B41073"
misc_feature	complement(36654. .37090)
	/note="match: GSS: Em:AQ772584"
misc_feature	37102. .37591
	/note="match: GSS: Em:AQ489567"
misc_feature	complement(37581. .38221)
	/note="match: GSS: Em:AQ418233"
repeat_region	37720. .37772

```
                /note="L2 repeat: matches 2657. .2710 of consensus"
     misc_feature    complement(37778. .38215)
                /note="match: GSS: Em:AQ819773"
     misc_feature    38218. .38732
                /note="match: GSS: Em:AQ599868"
     repeat_region   38861. .39164
                /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   39612. .39653
                /note="21 copies 2 mer ac 90% conserved"
     repeat_region   40854. .40909
                /note="L2 repeat: matches 2655. .2710 of consensus"
     repeat_region   42169. .42270
                /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   42411. .42548
                /note="MER5B repeat: matches 21. .181 of consensus"
     repeat_region   42780. .42938
                /note="MER5B repeat: matches 17. .177 of consensus"
     repeat_region   43229. .43326
                /note="MIR repeat: matches 40. .148 of consensus"
     repeat_region   45383. .45428
                /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                /note="match: STS: Em:G51995
                match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
                /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match             99.7%;  Score 2814;  DB 9;  Length 80247;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2817;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 AAGCTTTGCTACTCAGAGAGTCTCATATGCTTTAAGAAAAACTACATAGCAAAAGAAGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34271 AAGCTTTGCTACTCAGAGAGTCTCATATGCTTTAAGAAAAACTACATAGCAAAAGAAGGT 34330

Qy     61 GCTTTTTTTAATGTTCACATATATTTAGGTAATTAAATGAAGGATAATTTCTCCCGTTTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34331 GCTTTTTTTAATGTTCACATATATTTAGGTAATTAAATGAAGGATAATTTCTCCCGTTTT 34390

Qy    121 TTGGAAGGAAACATCTAATGTTTCTATGCTAGTTTCTGTCAAAATTTCAAAATACTTTAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34391 TTGGAAGGAAACATCTAATGTTTCTATGCTAGTTTCTGTCAAAATTTCAAAATACTTTAA 34450

Qy    181 TGAGTCTGAAATTTATAGGCCATATTATTTTCGTAAGAGAAGCAACATGAATTTCAAATG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34451 TGAGTCTGAAATTTATAGGCCATATTATTTTCGTAAGAGAAGCAACATGAATTTCAAATG 34510

Qy    241 ACCCGTGAGAATTTTATCAATAAACAGTTTTTGCAAATATATTTTTGTGTTCATAGCCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34511 ACCCGTGAGAATTTTATCAATAAACAGTTTTTGCAAATATATTTTTGTGTTCATAGCCAT 34570

Qy    301 AAAGAGGCATGAATGCCGAAGAGAAAAAAAACTTGAAAACACTTTTTTCTTTAAAAAAAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34571 AAAGAGGCATGAATGCCGAAGAGAAAAAAAACTTGAAAACACTTTTTTCTTTAAAAAAAAA 34630

Qy    361 GCCTTCACTCATCTTTTTAGCCACAATCTGCTAAAATTCATGACATCAAAGTTGTTCTTC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34631 GCCTTCACTCATCTTTTTAGCCACAATCTGCTAAAATTCATGACATCAAAGTTGTTCTTC 34690

Qy    421 ATGAAATATAGCCCTTTTTTGAGAATTTGTTTATTTTTATAATACCGATATCATCTTAAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34691 ATGAAATATAGCCCTTTTTTGAGAATTTGTTTATTTTTATAATACCGATATCATCTTAAT 34750

Qy    481 ACAAAGAAAAACTTTTTTAAAGCCCTTAGGCTATGTAAGTGAGAACGTTTTTCCAATGCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34751 ACAAAGAAAAACTTTTTTAAAGCCCTTAGGCTATGTAAGTGAGAACGTTTTTCCAATGCA 34810

Qy    541 AATAGAACATTTGATTATTACATCTTGGTAGATATCGGATGTACTAACAAATTATTAAAA 600
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  34811 AATAGAACATTTGATTATTACATCTTGGTAGATATTGGATGTACTAACAAATTATTAAAA 34870

Qy    601 GTAAACAAAGAAAACTGTAAGAAAAATATATAAATAAAAACAAAACACTGTGAGTTTTGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34871 GTAAACAAAGAAAACTGTAAGAAAAATATATAAATAAAAACAAAACACTGTGAGTTTTGA 34930

Qy    661 ATTTATAAAATTTGCCTGAGTGTACTTGTTTTTTAAAAATGTATCAAATCAAGAGTTGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34931 ATTTATAAAATTTGCCTGAGTGTACTTGTTTTTTAAAAATGTATCAAATCAAGAGTTGCC 34990

Qy    721 ATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTATAACG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34991 ATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTATAACG 35050

Qy    781 AGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCATCTGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35051 AGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCATCTGA 35110

Qy    841 ACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACTGGGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35111 ACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACTGGGA 35170

Qy    901 GATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTAGCGA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35171 GATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTAGCGA 35230

Qy    961 TGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACTGTTAAGGATAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  35231 TGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACAGTTAAGGATAG 35290

Qy   1021 CATTGAAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTTTGGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35291 CATTGAAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTTTGGG 35350

Qy   1081 AAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAATGGTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35351 AAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAATGGTT 35410

Qy   1141 TGATAACACTGAAAAGGAAAATGCTGTTGATTACTACTTGTTTATCTCATATTTATGATT 1200
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  35411 TGATAACACTGAAAAGGAAAATGCAGTTGATTACTACTTGTTTATCTCATATTTATGATT 35470

Qy   1201 TTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCCCTTC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35471 TTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCCCTTC 35530

Qy   1261 TGTCTTTCTCTATTTCGTAGGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35531 TGTCTTTCTCTATTTCGTAGGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCC 35590

Qy   1321 ATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35591 ATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAA 35650

Qy   1381 ACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35651 ACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAA 35710

Qy   1441 AAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35711 AAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTG 35770

Qy   1501 GCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35771 GCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATG 35830

Qy   1561 CCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35831 CCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCG 35890
```

```
Qy   1621 AGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35891 AGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATG 35950

Qy   1681 GATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35951 GATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTT 36010

Qy   1741 GTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36011 GTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTT 36070

Qy   1801 GACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36071 GACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTG 36130

Qy   1861 GGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36131 GGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAA 36190

Qy   1921 AAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36191 AAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTT 36250

Qy   1981 GATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGA 2040
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  36251 GATTCCTACCATATCTACTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGA 36310

Qy   2041 TTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36311 TTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGG 36370

Qy   2101 ACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGA 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36371 ACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGA 36430

Qy   2161 TTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36431 TTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT 36490

Qy   2221 ATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36491 ATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGA 36550

Qy   2281 ATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36551 ATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGG 36610

Qy   2341 TAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATAT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36611 TAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATAT 36670

Qy   2401 TTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36671 TTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTG 36730

Qy   2461 TCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36731 TCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTA 36790

Qy   2521 AAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTA 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36791 AAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTA 36850

Qy   2581 TGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTT 2640
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  36851 TGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGTTTTTTCACATTT 36910

Qy   2641 TTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36911 TTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGT 36970

Qy   2701 TTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTAT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36971 TTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTAT 37030

Qy   2761 TACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAAT 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37031 TACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAAT 37090

Qy   2821 TC 2822
          ||
Db  37091 TC 37092
```

RESULT 2
AB023021
LOCUS AB023021 3019 bp mRNA linear PRI 23-JUN-1999
DEFINITION Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX, complete cds.
ACCESSION AB023021
VERSION AB023021.1 GI:5139692
KEYWORDS alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S., Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and Narimatsu,H.
TITLE Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX
JOURNAL FEBS Lett. 453, 237-242 (1999)
REFERENCE 2 (bases 1 to 3019)
AUTHORS Kaneko,M., Kudo,T. and Narimatsu,H.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science, Soka University, Division of Cell Biology; Soka University, 1-236 Tangi-cho, Hachioji, Tokyo 192-8577, Japan
(E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132), Fax:81-426-91-9315)
FEATURES Location/Qualifiers
source 1. .3019
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 295. .1374
/gene="FUT9"
CDS 295. .1374
/gene="FUT9"
/codon_start=1
/product="alpha-1,3-fucosyltransferase IX"
/protein_id="BAA81685.1"
/db_xref="GI:5139693"
/translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT 937 a 594 c 541 g 947 t
ORIGIN

Query Match 54.6%; Score 1542; DB 9; Length 3019;
Best Local Similarity 100.0%; Pred. No. 2.4e-261;
Matches 1542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 346

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 406
```

```
Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 466

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 526

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 586

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    587 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 646

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    647 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 706

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    707 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 766

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    767 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 826

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    827 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 886

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    887 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 946

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    947 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 1006

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1007 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1066

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1067 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1126

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1127 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1186

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1187 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1246

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1247 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1306

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1307 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1366

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1367 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 1426

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1427 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 1486
```

```
Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1487 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 1546

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1547 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 1606

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1607 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 1666

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1667 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 1726

Qy   2721 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 2780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1727 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 1786

Qy   2781 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 2822
          |||||||||||||||||||||||||||||||||||||||||
Db    1787 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 1828
```

```
RESULT 3
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
```

```
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN
```

Query Match 54.5%; Score 1537.2; DB 9; Length 2501;
Best Local Similarity 99.8%; Pred. No. 1.7e-260;
Matches 1539; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 156

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 216

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 276

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 336

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 396

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    397 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGCTGGGATCTGA 456

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
          | || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 CTAACTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 516

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 576

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    577 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 636

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    637 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 696

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    697 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 756

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    757 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 816

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    817 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 876

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    877 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 936
```

```
Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 996

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    997 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1056

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1116

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1117 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1176

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 1236

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1237 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 1296

Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1297 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 1356

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1357 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 1416

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1417 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 1476

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1477 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 1536

Qy   2721 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 2780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1537 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 1596

Qy   2781 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 2822
          |||||||||||||||||||||||||||||||||||||||||
Db   1597 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 1638
```

```
RESULT 4
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
```

Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461, USA

```
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN
```

Query Match 39.1%; Score 1103.2; DB 10; Length 1705;
Best Local Similarity 85.4%; Pred. No. 3e-184;
Matches 1301; Conservative 0; Mismatches 213; Indels 10; Gaps 6;

```
Qy   1282 AAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTAT 1341
          ||||||||||||||||||||||||||| |||||||| |||||  |||| ||||  ||||
Db     55 AAAAATTATGACATCAACATCCAAAGGCATTCTTCGTCCATTCCTAATCGTCTGTGTTAT 114

Qy   1342 CCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTT 1401
          |||||  ||| |||||||||||||| || |||||||||| || |||||||||||  | ||
Db     115 CCTGGCCTGCTTCATGGCATGTCTGCTTATTTACATCAAGCCCACCAACAGCTGGGTGTT 174

Qy   1402 CAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAAC 1461
          ||||||||||| || ||||| |||||||||||||||||||| ||||| |||||  |||||
Db     175 CAGTCCAATGGAGTCTGCCAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAAC 234

Qy   1462 TGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGA 1521
          |||||||||||||||| |||||||| || ||||||||||||||||||||||||||||||||
Db     235 TGATTATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAGACCTTTGA 294

Qy   1522 CCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTC 1581
          |||||||||||||||| |||||||||||||||| |||||||||| || ||||| ||
Db     295 CCTTACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACAGACCGCTC 354

Qy   1582 ACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGAC 1641
          |||||||||||||||||||| || ||||| || ||| |||||||||| ||||||||||
Db     355 ACTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGAC 414

Qy   1642 AAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATC 1701
          || |||||||||| |||||||| ||||| ||||||||||||||||||  | || ||
Db     415 TAACTTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAACCTAGAGTC 474

Qy   1702 ACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTA 1761
          || ||||| || |||||||||||||||||||||||| ||||| |||||||||| || ||
Db     475 TCCCACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACTCTAACATA 534

Qy   1762 CCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTT 1821
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     535 TCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTT 594

Qy   1822 CGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCC 1881
          |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db     595 TGTGTTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAACTGGAATCC 654

Qy   1882 TGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTA 1941
          |||||||||||  |||||||||| |||||||| ||||| || ||||||||||||||||||
Db     655 GGAGCATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATCCATACCTA 714

Qy   1942 CGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTG 2001
          |||||||| |||||||||||| || ||||| |||||| | ||||| ||||||||| ||||
Db     715 TGGGCAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATATCTACTTG 774

Qy   2002 TAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATA 2061
          |||||||||||||||| |||||||| ||||| ||||| |||||||||| |||||||| ||
Db     775 TAAATTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAAAAGCTCTA 834

Qy   2062 CAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGA 2121
          |||||| ||| |||||||||| ||||||||||||||||||||| ||||||||||||||
Db     835 CAATGCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAAAACTATGA 894

Qy   2122 GAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCT 2181
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||| |
Db     895 GAATTATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGTT 954

Qy   2182 AGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 2241
          ||||| |||||  |||||||| |||||||||||||||| |||||||||| || |||||
Db     955 GGCAAAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTATTTCAACTG 1014

Qy   2242 GAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTG 2301
          |||||||||||||||||||| || ||||||||||||||||||||||||  |||||||
Db     1015 GAGGAAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGCCTGGCTTG 1074

Qy   2302 CGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTG 2361
          ||||| || |||||||| ||||| |||||||||| ||||||||||||||||||||||||
Db     1075 TGATCACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAATGGTTTTG 1134

Qy   2362 GAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAG 2421
          |||||||  |||  |    ||  |  |    |||| | || |   || |||||  |||
Db     1135 GAATTAATGTTTCCCTCATTTTCAAATTTAAGGAAATGTATTAACATGACATCATATAAG 1194

Qy   2422 TATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACC 2481
          |  | |||||||||||||   ||||   | ||||    ||||| |||||||| ||  ||
Db     1195 TTGTAAGGATAAGAAGAGA---AACACCTTGCTTTCAAATCACAGTTTATTTTAATTGCC 1251

Qy   2482 CTCTCTAGGGTAACGTGTATATTTTGGTGG-AGATTTTTAAAAGCTCAGCATGAGCAATC 2540
          ||| || | ||||| |||||||||| ||| | |||||||| ||||||| || |||||||
Db     1252 CTCCCTTGAGTAACATGTATATTTTGATGGAAAATTTTTAAGAGCTCAGAATTAGCAATC 1311

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
          | | |||| ||  |||||||||||||||||||| | || ||| ||||| || | ||||||
Db     1312 ACTTCATTTGG-ATTAAATTATCCTGTATATACGTGATAATGAGCACT-GAAAATAATTT 1369

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
          ||||| || | |||||||||||||  |||||||| | || |||||||  ||| |||
Db     1370 ATTCTCCACTCTCATTTGTAAACATTCTTTTTTCACA-TCTTATAGTTGTCTGTAAAGTA 1428

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
          || || || ||     |||||||| |||   ||||| | |||| ||| ||||||||||
Db     1429 AGTTTATGATT---CAATTGTTTCTACATCAATCAGATCTTTAGTCTGTTTGGGAAATGG 1485

Qy   2721 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 2780
          | |||    || |||| |||||||  ||||  | |||||||| |||||||| | | | |
Db     1486 AAATGGGTGTCCTAAAATATGAAAGCTTTTTGCAAAGTATTATAATGTCTAGATTCTAGT 1545

Qy   2781 TTGCATACTATAACAGAGGAAGAA 2804
          ||||||| |   |  | |||| |
Db     1546 TTGCATAATGGTAACAAAGAAGTA 1569
```

```
RESULT 5
AB015426
LOCUS       AB015426               2139 bp    mRNA    linear   ROD 23-JUN-1999
DEFINITION  Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete
            cds.
ACCESSION   AB015426
VERSION     AB015426.1  GI:3702718
KEYWORDS    Fut9; alpha1,3-fucosyltransferase IX.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K.
            and Narimatsu,H.
  TITLE     Expression cloning and characterization of a novel murine alpha1,
            3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15)
            epitope in brain and kidney
  JOURNAL   J. Biol. Chem. 273 (41), 26729-26738 (1998)
  MEDLINE   98434588
REFERENCE   2  (bases 1 to 2139)
  AUTHORS   Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN
```

Query Match 38.7%; Score 1092.2; DB 10; Length 2139;
Best Local Similarity 85.8%; Pred. No. 2.5e-182;
Matches 1309; Conservative 0; Mismatches 198; Indels 18; Gaps 8;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          |||||||||||||||||||||||||||||||| ||||||||||||||| |||| |||||||| |
Db    107 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 166

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| |||||||||||||| |||||||| ||||| || ||||||||||||| |||
Db    167 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 226

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||||| || || || ||||||||||||||||||| ||||| ||||| ||||
Db    227 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 286

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          |||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||||
Db    287 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 346

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          |||||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |
Db    347 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 406

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
          || |||||||||||||||||| || ||||| || ||  |||||||||| ||||||||||
Db    407 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 466

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
          | || |||||||||||  || |||||||||| ||||||||||||||||||||||||| || |
Db    467 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 526

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
          |||| |||||||| |||||||||||||||||||| |||||||| |||||||||| ||||
Db    527 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 586

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
          | |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    587 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 646

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
          | |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Db    647 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 706

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
          |||||||||||||| |||||||||||||| ||||| ||||| || ||||||||||| ||||
Db    707 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 766

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
          | || |||||||| |||||||| || || |||||||| ||||||| ||||||||| |||
Db    767 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 826

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
          |||||||||||||||| |||||||| ||||| ||||| |||||||||| |||||||| |
Db    827 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 886

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
          ||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||
Db    887 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 946

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
          |||||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||
Db    947 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 1006

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
          ||||||| |||||||||||||| |||||||||||||||| |||||||||||||||||||||
Db    1007 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 1066

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
          |||| |||||||| |||||||| || ||||| ||||||||||||||||||||  |||| |
Db    1067 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1126

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
          |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1127 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1186

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420
          |||||| ||| | | ||||| |  ||||   |  ||| |  || |||||| || || |||
Db    1187 GGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAA 1242

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480
          || || |||||||||||      ||||  || | |||| | |||  |||| ||||    |
Db    1243 GTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGC 1297

Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540
          |||| || | | | | |||||||||||||| | |||||||| || |||| || |||||||
Db    1298 CCTCCCTTGAGGACCATGTATATTTTGGTGAA-ATTTTTAAGAGATCAGAATTAGCAATC 1356

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
          | || ||| |||||||||||||||||||||  | || | | |||||||| | ||||||
Db    1357 ACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTT 1415

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
          |||  ||| | |||||| |||||||| |||| ||   |||| |||||| |  ||| |||
Db    1416 ATTTGTCACTCTCATTTATAAACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTA 1472

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
          |  || || ||  |  ||||||| |||  ||||| | ||||||||||||||| || |
Db    1473 AATTTATGATT---TACTTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAA 1529

Qy   2721 AGATGCACATCTTAAAGTATGA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAAC 2779
          | || || ||| |||| |||||  | ||||  | |||||||| || || ||| | | |
Db    1530 AAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAG 1589

Qy   2780 TTTGCATACTATAACAGAGGAAGAA 2804
          ||| ||| |  |  | ||||||
```

```
Db   1590 CTTGTATAATGCTAACAAAGAAGAA 1614


RESULT 6
AB049819
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN

  Query Match           38.5%;  Score 1085.4;  DB 10;  Length 2156;
  Best Local Similarity  84.3%;  Pred. No. 3.9e-181;
  Matches 1285;  Conservative    0;  Mismatches  226;  Indels   14;  Gaps    5;

Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          |||||||||||||| |||||||||||||| |||||||||||||| |||| ||||||||||
Db    338 GAAAAATTATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTA 397

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| ||| |||||||||| || || |||||||| || ||||||||||| |||
Db    398 TCCTGGGCTGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCT 457

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||||| || || |||||||||||||||||||||| ||||| ||||  ||||
Db    458 TCAGTCCAATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAA 517

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    518 CTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTG 577

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          |||||||||||||||||| |||||||||||||||| |||||||||||||||||||| |
Db    578 ACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCT 637

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
          | |||||||||||||||||| || ||||| || ||| |||||||||| ||||||||||
Db    638 CGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGA 697

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
          | || |||||||||| || ||||| ||||| |||||||||||||||||||| || || |
Db    698 CTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGT 757

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
          |||| |||||||| |||||||||||||||||||||||||||| |||||||||| ||||
Db    758 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTT 817

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
          | |||||||||||||||| |||||||||| |||||||||||||| ||||||| ||||
Db    818 ATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCT 877

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
          | |||||||||||||||||||||||||||| || ||||||||||||||||||||||||| |
Db    878 TTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATC 937

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
          | |||||||||||| |||||||||||||| ||||| ||||| || |||||||||| ||||
Db    938 CCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 997

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
          | |||||||||||||||||||||||| |||||||||||| ||||||| |||||||| |||
Db    998 ATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTT 1057

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
          |||||||||||||||||||||||||| ||||| ||||| |||||||||| |||||||| |
Db    1058 GTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 1117

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
          ||||||| ||| ||||||| || |||||||||| |||||||||||||||||||||||||
Db    1118 ACAATGCATTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAACTATG 1177

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
          |||||||||||||||| |||||||||||||||||||||||||| |||||| |||||||||
Db    1178 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGTGAGT 1237

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
          | ||||| |||||||||||| || |||||||||||||||||||||| |||||||||| ||||
Db    1238 TGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACT 1297

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
          |||||||||||||| |||||||| || ||||| |||||||||||||||||||  |||| |
Db    1298 GGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1357

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
          |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1358 GCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTT 1417

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420
          |||||||||| |     ||    |       |  ||||  || |||||  | ||| |||
Db    1418 GGAATTAAAATGT----CCATCATCGTCACAGTGAAATAGATTAATGAGACACCATTCAA 1473

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480
          || || |||||||||||      ||||  || | |||| |||||| |||||||||     |
Db    1474 GTGTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGTCACAGTTTATTTTTCCTGC 1528

Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540
          ||| || | | | | |||||||||||| || |  |||| | |  |           ||||
Db    1529 TCTCCCTTGAGGACCATGTATATTTTGATGAAATTTTTAAGAGACCAGAATCAGCAAATC 1588

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
          | || ||| |||||||||||||||||||||  | || |  |||||||| | ||||||
Db    1589 ACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACAAGCACTGGA-AATAATTT 1647
```

```
Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
          |||| ||||| |||||||| ||    | || |||     |||| |||||| |  ||| |||
Db   1648 ATTCATCATTCTCATTTGTGAAGCATTCTGTTTTTACATTTTATAGTTGACTGTAAAGTA 1707

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
          || ||||| ||   | ||||||||||||   ||||| | ||| ||||| ||||||||||
Db   1708 AGTTTGTGATT---TAATTGTTTCCACATCAATCAGATCTTTCATCTACTTGGGAAATGA 1764

Qy   2721 AGATGCACATCTTAAAGTATGAA-AAATTTTCACTAAGTATTACAATGTCTAGTTCCAAC 2779
          | || || ||  |||| ||||||  | | ||  | |||| ||| |  | ||| | | |
Db   1765 AAATACATATTCTAAAATATGAAGGACTCTTTGCCAAGTGTTATGAAATTTAGATTCTAA 1824

Qy   2780 TTTGCATACTATAACAGAGGAAGAA 2804
           ||| ||| |   |   |  ||| |
Db   1825 CTTGTATAATGCTAACAAAAAAGGA 1849


RESULT 7
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN

  Query Match            33.2%;  Score 937.6;  DB 10;  Length 1128;
  Best Local Similarity  91.4%;  Pred. No. 3.9e-155;
  Matches  994;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          |||||||||||| |||||||||||| |||||||||||| |||| ||||||||||
Db     41 GAAAAATTATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTA 100

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| ||| |||||||||| || || |||||||| || ||||||||||| |||
Db    101 TCCTGGGCTGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCT 160

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||| || || |||||||||||||||||||||| ||||| |||| ||||
Db    161 TCAGTCCAATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAA 220

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    221 CTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTG 280

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          |||||||||||||||||| |||||||||||||||| |||||||||||||||||||||| |
Db    281 ACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCT 340

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
          | |||||||||||||||||| || ||||| || ||| |||||||||| ||||||||||
Db    341 CGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGA 400

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
          | || |||||||||||| || ||||| ||||| ||||||||||||||||||||| || || |
Db    401 CTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGT 460

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
          |||| |||||||| |||||||||||||||||||||||||||| |||||||||| ||||
Db    461 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTT 520

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
          | |||||||||||||||| |||||||||| |||||||||||||| ||||||| ||||
Db    521 ATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCT 580

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
          | |||||||||||||||||||||||||| || ||||||||||||||||||||||||| |
Db    581 TTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATC 640

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
          | |||||||||| |||||||||||| ||||| ||||| || |||||||||| ||||
Db    641 CCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 700

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
          | |||||||||||||||||||| |||||||||||| ||||||| |||||||| |||
Db    701 ATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTT 760

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
          |||||||||||||||||||||||| ||||| ||||| |||||||||| |||||||| |
Db    761 GTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 820

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
          ||||||| ||| ||||||| |||||||||||||| ||||| ||||||||||||||||||
Db    821 ACAATGCATTTTTGGCTGGTTCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 880

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
          |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db    881 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 940

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
          | ||||| |||||||||| || |||||||||||||||||||| |||||||||| ||||
Db    941 TGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACT 1000

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
          |||||||||||| |||||||| || ||||| ||||||||||||||||||||  |||| |
Db   1001 GGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1060

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
          |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1061 GCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTT 1120

Qy   2361 GGAATTAA 2368
          ||||||||
Db   1121 GGAATTAA 1128
```

```
RESULT 8
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match           29.2%;  Score 823.2;  DB 5;  Length 1707;
  Best Local Similarity   76.4%;  Pred. No. 4.4e-135;
  Matches 1068;  Conservative    1;  Mismatches  309;  Indels   20;  Gaps    4;

Qy   1268 CTCTATTTCGTAGGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTA 1327
          | ||   | | || | |||||||||||||||||||| ||||||||| | || ||||||||
Db    247 CACTGCCTGGCAGAACAAATTATGACATCAACATCTAAAGGAATTTTCCGGCCATTTTTT 306

Qy   1328 ATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACC 1387
          ||||||| |||| |:|| || |||||||||||||| | | || ||||| || ||||| ||
Db    307 ATTGTCTTCATTGTYCTTGGTTGTTTCATGGCATTTATACTAATTTATATTAAACCAACA 366

Qy   1388 AACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTC 1447
          | ||||||||||  | |||  |||||||||||||||| ||  |||||||||| | ||||
Db    367 AGCAGCTGGATCTCTAATCCTGTGGAATCAGCCAGCTCAGTTTTGAAAATGAAGAGCTTC 426

Qy   1448 TTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTT 1507
          |||||  |||||||||||| |||||||||||||||| |||| |||| ||| | |||||||||
Db    427 TTTTCTTCCAAAACTGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTT 486

Qy   1508 GGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTC 1567
          || ||||| || || || |||||||||||| | |||||||||||||| || ||||||||
Db    487 GGCCAGACATTCGATCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTG 546

Qy   1568 ACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATC 1627
          || |  ||||| ||||| || ||||| |||||||||||||| || |||||| | |||||

Db    547 ACTATTGACCGCTCACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATT 606

Qy   1628 AGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGG 1687
          |  |||||||||| |||||||||||||||||| |||||||| |||||||||||||||||
Db    607 AACTGGGATCTGACTAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGG 666

Qy   1688 ATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAAC 1747
          ||||| |||||||| |||||||| ||||| |||||||||||||||||| ||| | ||||||
Db    667 ATGAACTTGGAATCTCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAAC 726

Qy   1748 CTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTA 1807
          || || |||||||||| |||||||||||||| || |||||||||||| | |||  ||
Db    727 CTAACCCTGACTTACCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTT 786

Qy   1808 AGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTG 1867
          | |||   ||||   | |||||||||||| ||||| || ||||| || ||||||||
Db    787 GGTACAGGCGCCTTTACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTA 846

Qy   1868 AGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATT 1927
          |||||||||||||||||||||  |||||||||||||||||||| |||||   ||||
Db    847 AGTAACTGGAACCCTGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATT 906

Qy   1928 GAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCT 1987
          |||||||||||||||||||| |||||||| || ||||| |||||||| || |||||
Db    907 GAAATCCATACCTACGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCA 966

Qy   1988 ACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATC 2047
          || || ||   ||| ||||| || ||||| |||||||||||||||||||| ||||||||
Db    967 ACTATCTCCAGTTGCAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATT 1026

Qy   2048 ACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCT 2107
          || || || || |||||||||| |||||||||| || ||||| ||||| |||||| || ||
Db    1027 ACTGAGAAACTTTACAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCC 1086

Qy   2108 AGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAAC 2167
          || |||||||| |||||||| |||||||||| || || || || |||||||||| |
Db    1087 AGAGAAAACTACGAGAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTA 1146

Qy   2168 TCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTT 2227
          || || || ||||| ||| | ||||| |     | |||||||| || ||| | ||||||
Db    1147 TCCCCAAGAGAGCTTGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTT 1206

Qy   2228 AGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACAT 2287
          |||||||| |||||||||||||||||| | ||  |||||||  ||| ||||||||  |||
Db    1207 AGTTACTTCAACTGGAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCAT 1266

Qy   2288 GCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTA 2347
          |||||| | ||||| |||||||||||||| || || ||||| |||||  |||| ||||||
Db    1267 GCATGTCTTGCTTGTGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTA 1326

Qy   2348 GAGAAATGGTTTTGGAATTAA--AATTTTTCATCACTTGCACACTTGATAAATATTTTGA 2405
          || |||||||||||||| ||     | | |     |  ||||| ||   ||  | |
Db    1327 GAAAAATGGTTTTGGAACTAATTTGTGTGTGTGTATGTGCACTTTTTTGAAGAAGATAAT 1386

Qy   2406 TGAGATATCATCCAAGTATTGAGGATAAGAA----------------GAGATGCAACATA 2449
          ||  | |  || ||||   ||| | | |||                ||    ||| |
Db    1387 GGATCTTTAGTCTAAGTGGAGAGTAAAGGAATAAGGAAAAAAGAAGGAAGGATGAACCTC 1446

Qy   2450 CTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGT 2509
          |       |    |||  | | |    || ||||||| || ||   | | ||| ||
Db    1447 ACAGAACCTTACGTCAAGGTTTATCAGCCATCCTCTCTTGGCTATTCTATTTATATTTTG 1506

Qy   2510 GGAGATTTTTAAAAGCTCAGCATGAGCAATCAT-TCCATTCGGTTTTAAATTATCCTGTA 2568
          |||  ||| |||  | |  |  |||| ||||     | || ||||||||||  ||||
Db    1507 TAAGAGATTTTAAAAATTATGAGCAGCAGTCATCAAAACTCAGTTTTAAATTATAATGTA 1566

Qy   2569 TATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGC 2628
          |||  |||||||||||||||| ||| ||  | ||| || | | ||| |||| ||||||
Db    1567 CATA-CTAATTATGTGCACTGAAGAATATATGATTGTTTACTATAATTTTTAAACAAGAT 1625

Qy   2629 TTTTTCACATTTTTGTAG 2646
          ||||     ||| ||| |
Db    1626 TTTTCACATTTTCTGTGG 1643
```

```
RESULT 9
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus`laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN

  Query Match             24.0%;  Score 676.8;  DB 5;  Length 1080;
  Best Local Similarity   76.7%;  Pred. No. 2.5e-109;
  Matches  828;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy    1289 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 1348
           ||||||||| |||| ||||||||  | ||  ||||||||  |||| ||  | |  ||  |
Db       1 ATGACATCAGCATCTAAAGGAATCTTACGGGCATTTTTAGCTGTCAGCGGTGTGTTGATC 60

Qy    1349 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 1408
           |||||  |||  |||| | || || ||| ||||||| || |||| ||||||  | ||||||
Db      61 TGTTTTATGATATGTTTACTGATATACGTCAAACCAACAAACAACTGGATTTCCAGTCCT 120

Qy    1409 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 1468
           || ||||||||     || || || |||||||| || || || ||| | |||    |||||
Db     121 ATCGAATCAGCACATTCAGTCCTTAAAATGAAGAATTTTTTCTCCTCTAAACAGAATTAT 180

Qy    1469 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 1528
           |  ||||||||     ||||  | ||||||||||| |||||||||| ||||| || | |
Db     181 TACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGACTTTTGAACTGAAA 240

Qy    1529 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 1588
           ||||||||  || |||| || ||  | || |||||||| || || |||||  |||| |||
Db     241 TCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTGACCGCGCACTCTAC 300

Qy    1589 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 1648
           |||||  || |||||  || ||  | || |||||||| || || |||||| ||||| |||||
Db     301 AACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGGATCTCATTAACCTG 360

Qy    1649 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 1708
           || || ||    |||||||| |||||||||||||||||||||||| ||||||||||||
Db     361 CCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATCTGGAATCACCAACA 420

Qy    1709 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 1768
           || || || ||||| |||||||||||||||| |||||||||||||||||||| | |||
Db     421 CATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTCTGACTTACAGGCGT 480

Qy    1769 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
           || |||||||||||||||||||||||||||| ||    |||||||||| || || |  |||
Db     481 GACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAAAACCATTTGACTTT 540

Qy    1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
           || |||||||||||||| || || || || |||||||| || |||||||||| || |||
Db     541 GAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACTGGAACCCAGAACAT 600

Qy    1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
           ||||||||||| |||||||||| || | |||| |||||||||    || || ||||||
Db     601 GCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTATAACTTATGGGCAA 660

Qy    1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
           || ||||| |||||  | |  |||||||   || |||| || ||  |  |||| ||||||
Db     661 GCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTGCATCTTGCAAATTT 720

Qy    2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 2068
           ||||| || |||||||||| || ||||| ||||| ||||| ||||| || || |||||
Db     721 TATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAAAACTTTATAATGCA 780

Qy    2069 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 2128
           ||  |||||| || || ||  |||| |||||||  |||| ||||||||||||| ||
Db     781 TTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAAACTATGAGAACTAC 840

Qy    2129 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 2188
           ||||| |||||||| || || || |||||||||| |  |||||| || ||  ||||  |
Db     841 ATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAAGAGAATTAGCTGAC 900

Qy    2189 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
           ||||        |  ||||| | |   |    ||||| || ||||  ||||||||||
Db     901 CATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACTACAACTGGAGGAAA 960

Qy    2249 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 2308
           |||| || || ||| | || |  || |||||||| |||||||| || || || || |||
Db     961 CATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCTTAGCGTGTGACCAT 1020

Qy    2309 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 2368
           ||||||||||||||||||| || || || || ||| |||||| ||||| ||||| |||
Db    1021 GTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGTGGTTCTGGAACTAA 1080

RESULT 10
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
```

Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L15932
Center clone name: 167_K_24
----------.
* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 788: contig of 788 bp in length
* 789 888: gap of 100 bp
* 889 1702: contig of 814 bp in length
* 1703 1802: gap of 100 bp
* 1803 2642: contig of 840 bp in length
* 2643 2742: gap of 100 bp
* 2743 3637: contig of 895 bp in length
* 3638 3737: gap of 100 bp
* 3738 4635: contig of 898 bp in length
* 4636 4735: gap of 100 bp
* 4736 5657: contig of 922 bp in length
* 5658 5757: gap of 100 bp
* 5758 6692: contig of 935 bp in length
* 6693 6792: gap of 100 bp
* 6793 7703: contig of 911 bp in length
* 7704 7803: gap of 100 bp
* 7804 8742: contig of 939 bp in length
* 8743 8842: gap of 100 bp
* 8843 9787: contig of 945 bp in length
* 9788 9887: gap of 100 bp
* 9888 10819: contig of 932 bp in length
* 10820 10919: gap of 100 bp
* 10920 11808: contig of 889 bp in length
* 11809 11908: gap of 100 bp
* 11909 12830: contig of 922 bp in length
* 12831 12930: gap of 100 bp
* 12931 13847: contig of 917 bp in length
* 13848 13947: gap of 100 bp
* 13948 14886: contig of 939 bp in length
* 14887 14986: gap of 100 bp
* 14987 15911: contig of 925 bp in length
* 15912 16011: gap of 100 bp
* 16012 16903: contig of 892 bp in length
* 16904 17003: gap of 100 bp
* 17004 17906: contig of 903 bp in length
* 17907 18006: gap of 100 bp
* 18007 18931: contig of 925 bp in length
* 18932 19031: gap of 100 bp
* 19032 19945: contig of 914 bp in length
* 19946 20045: gap of 100 bp
* 20046 20951: contig of 906 bp in length
* 20952 21051: gap of 100 bp
* 21052 21949: contig of 898 bp in length
* 21950 22049: gap of 100 bp
* 22050 22950: contig of 901 bp in length
* 22951 23050: gap of 100 bp
* 23051 23945: contig of 895 bp in length
* 23946 24045: gap of 100 bp
* 24046 24955: contig of 910 bp in length
* 24956 25055: gap of 100 bp
* 25056 25980: contig of 925 bp in length
* 25981 26080: gap of 100 bp
* 26081 27052: contig of 972 bp in length
* 27053 27152: gap of 100 bp
* 27153 28067: contig of 915 bp in length
* 28068 28167: gap of 100 bp
* 28168 29076: contig of 909 bp in length
* 29077 29176: gap of 100 bp
* 29177 30093: contig of 917 bp in length
* 30094 30193: gap of 100 bp
* 30194 31099: contig of 906 bp in length
* 31100 31199: gap of 100 bp
* 31200 32097: contig of 898 bp in length
* 32098 32197: gap of 100 bp
* 32198 33113: contig of 916 bp in length
* 33114 33213: gap of 100 bp
* 33214 34125: contig of 912 bp in length
* 34126 34225: gap of 100 bp
* 34226 35133: contig of 908 bp in length
* 35134 35233: gap of 100 bp
* 35234 36150: contig of 917 bp in length
* 36151 36250: gap of 100 bp
* 36251 37171: contig of 921 bp in length
* 37172 37271: gap of 100 bp
* 37272 38159: contig of 888 bp in length
* 38160 38259: gap of 100 bp
* 38260 39156: contig of 897 bp in length
* 39157 39256: gap of 100 bp
* 39257 40188: contig of 932 bp in length
* 40189 40288: gap of 100 bp
* 40289 41210: contig of 922 bp in length
* 41211 41310: gap of 100 bp
* 41311 42222: contig of 912 bp in length
* 42223 42322: gap of 100 bp
* 42323 43218: contig of 896 bp in length
* 43219 43318: gap of 100 bp
* 43319 44233: contig of 915 bp in length
* 44234 44333: gap of 100 bp
* 44334 45246: contig of 913 bp in length
* 45247 45346: gap of 100 bp
* 45347 46261: contig of 915 bp in length
* 46262 46361: gap of 100 bp
* 46362 47279: contig of 918 bp in length
* 47280 47379: gap of 100 bp
* 47380 48276: contig of 897 bp in length
* 48277 48376: gap of 100 bp
* 48377 49282: contig of 906 bp in length
* 49283 49382: gap of 100 bp
* 49383 50305: contig of 923 bp in length
* 50306 50405: gap of 100 bp
* 50406 51314: contig of 909 bp in length
* 51315 51414: gap of 100 bp
* 51415 52335: contig of 921 bp in length

```
            *    52336 52435: gap of      100 bp
            *    52436    53163: contig of 728 bp in length
            *    53164 53263: gap of      100 bp
            *    53264    54210: contig of 947 bp in length
            *    54211 54310: gap of      100 bp
            *    54311    55199: contig of 889 bp in length
            *    55200 55299: gap of      100 bp
            *    55300    56212: contig of 913 bp in length
            *    56213 56312: gap of      100 bp
            *    56313    57287: contig of 975 bp in length
            *    57288 57387: gap of      100 bp
            *    57388    58245: contig of 858 bp in length
            *    58246 58345: gap of      100 bp
            *    58346    59236: contig of 891 bp in length
            *    59237 59336: gap of      100 bp
            *    59337    60272: contig of 936 bp in length
            *    60273 60372: gap of      100 bp
            *    60373    61289: contig of 917 bp in length
            *    61290 61389: gap of      100 bp
            *    61390    62299: contig of 910 bp in length
            *    62300 62399: gap of      100 bp
            *    62400    63322: contig of 923 bp in length
            *    63323 63422: gap of      100 bp
            *    63423    64336: contig of 914 bp in length
            *    64337 64436: gap of      100 bp
            *    64437    65354: contig of 918 bp in length
            *    65355 65454: gap of      100 bp
            *    65455    66352: contig of 898 bp in length
            *    66353 66452: gap of      100 bp
            *    66453    67362: contig of 910 bp in length
            *    67363 67462: gap of      100 bp
            *    67463    68370: contig of 908 bp in length
            *    68371 68470: gap of      100 bp
            *    68471    69379: contig of 909 bp in length
            *    69380 69479: gap of      100 bp

 Query Match             22.8%;  Score 644;  DB 2;  Length 71396;
 Best Local Similarity   75.1%;  Pred. No. 6.7e-104;
 Matches  784;  Conservative    0;  Mismatches  247;  Indels   13;  Gaps    4;

Qy   1215 AATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCCCTTCTGTCTTTCTCTATT 1274
          || | | |  || | |    || | |   |  |  |  | |||
Db  65311 AAAACAGTAGTGGTATGGAATTTACTTTAAATGAAATATACCCANNNNNNNNNNNNNNNN 65370

Qy   1275 TCGTAGGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCT 1334

Db  65371 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 65430

Qy   1335 GCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCT 1394
                                  || |||   | | |     |  || ||||||||||
Db  65431 NNNNNNNNNNNNNNNNNNNNNNNGATCTCTCGCCCTGTGGTGGATTCCCACCAACAGCT 65490

Qy   1395 GGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCA 1454
           | ||||||||||||||||| || || || |||||||||||||||||| ||||| ||||
Db  65491 CGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCA 65550

Qy   1455 CCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGA 1514
          | |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db  65551 CAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGA 65610

Qy   1515 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGG 1574
          |||||||||||||||||||||||||||||||||| |||||||| |||||||||||||| |
Db  65611 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAG 65670

Qy   1575 ACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGG 1634
          |||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| ||||
Db  65671 ACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGG 65730

Qy   1635 ATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATT 1694
          ||||||| || |||||||||||| || |||||||||| ||||||||||||||||||||||
Db  65731 ATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATT 65790

Qy   1695 TGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTC 1754
          | || ||||| |||||||| |||||||||||||||||| |||||||| ||||||||||
Db  65791 TAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTC 65850

Qy   1755 TGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAA 1814
          | ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65851 TAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAA 65910

Qy   1815 ATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 1874
          ||||||| ||||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db  65911 ATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACT 65970

Qy   1875 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 1934
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||||||
Db  65971 GGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCC 66030

Qy   1935 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 1994
          | ||||| || |||||||| |||||||| || || ||||||||| ||||||| |||||||
Db  66031 ACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATAT 66090

Qy   1995 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 2054
          || |||||||||||||||||  |  ||| |    ||  |  | | |   | |
Db  66091 CTACTTGTAAATTTTATCTTTCTCTGAAAACTCATTCACAAGATTACTCCCAAAAG---- 66146

Qy   2055 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 2114
               | ||     ||| ||||||| || |||||||||| ||||| ||||||||||||
Db  66147 -----TCCACATGCTTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAA 66201

Qy   2115 ACTATGAGAATTATATTCCAGCAGATTCATTCATTC--ATGTGGAAGATTATAACTCT-C 2171
          |||||||||||||||||||||| ||||||||||||    |||||||||| | ||||| |
Db  66202 ACTATGAGAATTATATTCCAGCTGATTCATTCATTCCATCGTGGAAGATTTTCACTCTCC 66261

Qy   2172 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT-ATACCTTAGT 2230
          |||||||| |||||||    |||||||||||||||| ||||||||||  |||||||||
Db  66262 CCAGTGAGTTAGCAAATATCCTGAAGGAAGTCGACAANAACAATAAGTTGGTACCTTAGT 66321

Qy   2231 TACTTTAACTGGAGGAAGGATTTC 2254
          |||| ||||| |  ||| |  |||
Db  66322 TACTCTAACTCGNAGAAAGGATTC 66345
```

```
RESULT 11
G55571/c
LOCUS       G55571                   561 bp    DNA     linear   STS 30-MAR-2000
DEFINITION  SHGC-100926 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G55571
VERSION     G55571.1  GI:6120890
KEYWORDS    STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Olivier,M. and Cox,D.R.
  TITLE     Unpublished, Olivier, M., Cox, D.R. (2000)
  JOURNAL   Unpublished
COMMENT
            Contact: Michael Olivier, David R. Cox
            Stanford Human Genome Center
            Stanford University School of Medicine
            4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
            Tel: (650) 320-5800
            Fax: (650) 320-5801
            Email: olivier@shgc.stanford.edu
            Primer A: CAGAACTTTCCTCAAAGCCAAGA
            Primer B: TAACGAATTCCATCTGAACCACA
            STS size: 302
            PCR Profile:
                    Initial incubation: 95 degrees C for 10 minutes
                    Denaturation:       94 degrees C for 30 seconds
                    Annealing:          60 degrees C for 30 seconds
                    Polymerization:     72 degrees C for 23 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9700
```

```
          Protocol:
                 Template:                  25 ng
                 Primer:                    each 1 uM
                 dNTPs:                     each 200 uM
                 AmpliTaq Gold Polymerase:  0.07 units/ul
                 Total Vol:                 5 ul

          Buffer:
                 MgCl2:          2.5 mM
                 KCl:            50 mM
                 Tris-HCl:       10 mM
                 pH:             8.3

          BAC ends sequenced at TIGR from the RPCI11 BAC library. Designed
          and developed at the Stanford Human Genome Center.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6"
                     /clone_lib="Human"
     STS             153. .454
     primer_bind     153. .175
     primer_bind     complement(432. .454)
BASE COUNT      200 a     92 c     89 g    180 t
ORIGIN

  Query Match             19.8%;  Score 557.8;  DB 11;  Length 561;
  Best Local Similarity   99.6%;  Pred. No. 2.2e-88;
  Matches  559;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     717 TGCCATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTAT 776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 TGCCATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTAT 502

Qy     777 AACGAGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCAT 836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 AACGAGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCAT 442

Qy     837 CTGAACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACT 896
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 CTGAACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACT 382

Qy     897 GGGAGATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTA 956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 GGGAGATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTA 322

Qy     957 GCGATGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACTGTTAAGG 1016
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     321 GCGATGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACAGTTAAGG 262

Qy    1017 ATAGCATTGAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTT 1076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ATAGCATTGAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTT 202

Qy    1077 TGGGAAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAAT 1136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 TGGGAAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAAT 142

Qy    1137 GGTTTGATAACACTGAAAAGGAAAATGCTGTTGATTACTACTTGTTTATCTCATATTTAT 1196
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     141 GGTTTGATAACACTGAAAAGGAAAATGCAGTTGATTACTACTTGTTTATCTCATATTTAT 82

Qy    1197 GATTTTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCC 1256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GATTTTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCC 22

Qy    1257 CTTCTGTCTTTCTCTATTTCG 1277
           |||||||||||||||||||||
Db      21 CTTCTGTCTTTCTCTATTTCG 1

RESULT 12
```

```
AC100674/c
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      788: contig of 788 bp in length
            *     789 888: gap of      100 bp
            *     889     1702: contig of 814 bp in length
            *    1703 1802: gap of      100 bp
            *    1803     2642: contig of 840 bp in length
            *    2643 2742: gap of      100 bp
            *    2743     3637: contig of 895 bp in length
            *    3638 3737: gap of      100 bp
            *    3738     4635: contig of 898 bp in length
            *    4636 4735: gap of      100 bp
```

```
*     4736     5657: contig of 922 bp in length
*     5658 5757: gap of      100 bp
*     5758     6692: contig of 935 bp in length
*     6693 6792: gap of      100 bp
*     6793     7703: contig of 911 bp in length
*     7704 7803: gap of      100 bp
*     7804     8742: contig of 939 bp in length
*     8743 8842: gap of      100 bp
*     8843     9787: contig of 945 bp in length
*     9788 9887: gap of      100 bp
*     9888    10819: contig of 932 bp in length
*    10820 10919: gap of      100 bp
*    10920    11808: contig of 889 bp in length
*    11809 11908: gap of      100 bp
*    11909    12830: contig of 922 bp in length
*    12831 12930: gap of      100 bp
*    12931    13847: contig of 917 bp in length
*    13848 13947: gap of      100 bp
*    13948    14886: contig of 939 bp in length
*    14887 14986: gap of      100 bp
*    14987    15911: contig of 925 bp in length
*    15912 16011: gap of      100 bp
*    16012    16903: contig of 892 bp in length
*    16904 17003: gap of      100 bp
*    17004    17906: contig of 903 bp in length
*    17907 18006: gap of      100 bp
*    18007    18931: contig of 925 bp in length
*    18932 19031: gap of      100 bp
*    19032    19945: contig of 914 bp in length
*    19946 20045: gap of      100 bp
*    20046    20951: contig of 906 bp in length
*    20952 21051: gap of      100 bp
*    21052    21949: contig of 898 bp in length
*    21950 22049: gap of      100 bp
*    22050    22950: contig of 901 bp in length
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp

  Query Match             10.0%;  Score 281;  DB 2;  Length 71396;
  Best Local Similarity   65.4%;  Pred. No. 3.7e-40;
  Matches  465;  Conservative    0;  Mismatches  227;  Indels   19;  Gaps    9;

Qy    2096 CTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTG 2155
           | | | | |   |  |
Db   30210 CAGTGGCGAGACATCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 30151

Qy    2156 GAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAAT 2215
                                                                  |||
Db   30150 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAT 30091

Qy    2216 AAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCAC-GATT 2274
           ||||  ||||||||||||||||||||| |||||||| |||||||| || |||| | ||
Db   30090 AAGTNGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGGTT 30031

Qy    2275 TTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTC 2334
```

```
Db  30030 TTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTC 29971

Qy   2335 TGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGAT 2394

Db  29970 TGTTGGTAATTTAGAGAAATGGTTTTGGAATT-AAAGTGTCCATCATTGTCACA---GTG 29915

Qy   2395 AAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACT 2454

Db  29914 AAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA-----AACACGCTGCA 29860

Qy   2455 TTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGA 2514

Db  29859 TTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGTGAA-A 29801

Qy   2515 TTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACC 2574

Db  29800 TTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATG 29741

Qy   2575 TAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTC 2634

Db  29740 TGATAACGAGCACTGGA-AATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTAC 29682

Qy   2635 ACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGATC 2694

Db  29681 A---TTTTATAGTTGACTGTAAAGTAAATTTATGATT---TACTTGTTTCTACATCAATC 29628

Qy   2695 AGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGA-AAAATTTTCAC 2753

Db  29627 AGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTTTGC 29568

Qy   2754 TAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAA 2804

Db  29567 CAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAA 29517
```

```
RESULT 13
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN
```

Query Match 9.9%; Score 279.6; DB 5; Length 1567;
Best Local Similarity 58.3%; Pred. No. 1.3e-39;
Matches 533; Conservative 0; Mismatches 369; Indels 12; Gaps 2;

```
Qy   1471 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 1530

Db    461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy   1531 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 1590

Db    521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy   1591 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 1650

Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGG---------ATTTGCC 631

Qy   1651 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 1710

Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy   1711 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 1770

Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy   1771 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTT 1827

Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy   1828 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 1887

Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy   1888 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 1947

Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy   1948 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 2007

Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy   2008 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 2067

Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy   2068 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 2127

Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy   2128 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 2187

Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy   2188 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 2247

Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy   2248 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 2307

Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy   2308 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 2367
```

```
Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy   2368 AAATTTTTCATCAC 2381
          |     ||  ||||
Db   1352 ATGGGATTGTTCAC 1365


RESULT 14
AB023628
LOCUS       AB023628                1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match             8.3%;  Score 234.6;  DB 5;  Length 1832;
  Best Local Similarity   55.5%;  Pred. No. 1e-31;
  Matches  500;  Conservative    0;  Mismatches  389;  Indels   12;  Gaps    2;

Qy   1471 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 1530
          | | || |  |   |  || | ||| ||   || ||  |     | |||||||| | |
Db    619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy   1531 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 1590
           ||          ||  |||||||| || || ||  | |||   ||| |    |||  | |
Db    679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy   1591 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 1650
           ||| | |||| |||| || | || ||| |||| || ||          ||| |||| ||
Db    739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGC---------ACAGATTTTCC 789

Qy   1651 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 1710
           || |   |  | |||   ||||| || ||||| |||  |||| |||| || || || ||
Db    790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy   1711 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 1770
               |||||       | |   |   | ||||||  ||||     | ||||  | | ||
Db    850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy   1771 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA---AATCCCTTCGTGTT 1827
          |||||||||  | |||||||||  ||      |    |    ||||    | |       |||
Db    910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy   1828 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 1887
            | || |  || || ||| | || |||||| | || |||||||   |     | ||
Db    970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy   1888 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 1947
           |     | |  | | | ||| |    |||  |||| |  |    ||||| ||
Db    1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy   1948 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 2007
           ||      |      | |||| | |  |||        |  | ||||   |||||||||
Db    1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy   2008 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 2067
           || || |||||||| || || || |||| || ||  |||| || ||||| | |||| |
Db    1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy   2068 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 2127
          |||        ||  | || ||||| ||  | || ||  | ||  | |||||||| | |
Db    1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy   2128 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 2187
           || |||  |||  | ||||| |||||||| ||||    . ||||||  || || ||  |
Db    1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy   2188 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 2247
            | || |      | ||| |||||  |  |    |||       |  || ||||||||||
Db    1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy   2248 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 2307
            ||    |  |||   | ||   ||| ||      |||| | |||  |  ||| ||| |
Db    1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy   2308 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 2367
          | | |||||| || |  | ||||   | |||    || |  |    |||| ||| |||
Db    1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy   2368 A 2368
          |
Db    1510 A 1510


RESULT 15
AB035907
LOCUS       AB035907                1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
```

Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
Tel:81-559-81-6790, Fax:81-559-81-6789)

```
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN
```

Query Match 7.9%; Score 222.6; DB 5; Length 1092;
Best Local Similarity 57.3%; Pred. No. 1.4e-29;
Matches 484; Conservative 0; Mismatches 349; Indels 12; Gaps 4;

```
Qy   1472 AATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 1531
          ||  ||   ||| |  |  | |||     |||||||||  |           |||
Db    199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy   1532 TGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAAC 1591
          |||||     | |||||  ||   || |||||  |||||||  || |    || ||| |
Db    259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy   1592 AAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCT 1651
           |  |  | ||  || || |||| ||||| |||||   ||  ||| | |    |  | |||
Db    319 GAGGCAGACGCCATTATGTTCCACCACCGGGACAT---TTCTGATTTCAGTGACCTTCCT 375

Qy   1652 CAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 1711
             | |    | |    || ||| || ||||| |||||||| || ||||| ||  | ||
Db    376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy   1712 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 1771
           | |||     |    |  | | |  | | || |||  |||  || | ||| |    ||
Db    436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy   1772 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 1831
          || ||||||    |||| || ||||||||      |   ||  || |    |    | |
Db    496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy   1832 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 1891
           ||||  ||||    ||  |||||  ||||||  | || |||||||||   || || |
Db    553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy   1892 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 1951
          || ||| |||||||||| ||||| || ||    | | |||||| |    |||||  |
Db    613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy   1952 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 2011
          | |||     || | || ||  | ||  |  |    ||| | || |  | |||||||||
Db    670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy   2012 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 2068
           |  | ||||| |||||  | |||| |||||||||||||||||||| |     || ||
Db    730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy   2069 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 2128
          |||    |  | |  | ||  | ||  |||| ||     |  | ||||| ||||  |
Db    790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy   2129 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 2188
          || ||       || |||||||| ||||| ||||    |   |||||  |||| || | |
Db    850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy   2189 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          || |||||| |  | || |||||||| | | || ||||  || || || |  |||| ||||
Db    910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy   2249 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 2308
          || |   ||  |  |  |    || |||||    |||   ||    || ||| |
Db    970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy   2309 GTGAA 2313
          || ||
Db   1030 GTCAA 1034
```

Search completed: October 6, 2002, 04:30:09
Job time : 4349.27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 316.687 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-4
Perfect score: 2822
Sequence: 1 aagctttgctactcagagag.....aacatgttgcgattgaattc 2822

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2822	100.0	2822	21	AAZ92646	Human alpha-1,3-fu
2	1542	54.6	2676	21	AAZ92647	Human alpha-1,3-fu
3	1092.2	38.7	2036	21	AAZ92672	Murine alpha-1,3-f
4	1092.2	38.7	2170	21	AAZ92645	Murine alpha-1,3-f
5	127.4	4.5	1814	18	AAT59505	Murine myeloid-lin
6	127	4.5	3594	18	AAT85219	Mouse alpha-fucosy
7	114.4	4.1	1701	15	AAQ77732	Human alpha-1,3-fu
8	114.4	4.1	1701	21	AAF21136	Human low adenosin
9	114.4	4.1	1701	21	AAA35014	Human adenosine re
10	114.4	4.1	6944	21	AAF21137	Human low adenosin
11	114.4	4.1	6944	21	AAA35015	Human adenosine re
12	113.4	4.0	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
13	113.4	4.0	1654	18	AAT61680	Human alpha(1,3)-f
14	112.2	4.0	1086	15	AAQ56912	Lewis blood group
15	112.2	4.0	2042	18	AAT80111	Fucosyltransferase
16	112.2	4.0	2043	12	AAQ13330	Glycosyltransferas
17	112.2	4.0	2043	15	AAQ56906	DNA encoding a gly
18	112.2	4.0	2043	18	AAT76769	Human alpha 1,3/4
19	112.2	4.0	2043	18	AAT61675	Human alpha(1,3/1,
20	111.8	4.0	1126	21	AAF21135	Human low adenosin
21	111.8	4.0	1126	21	AAA35013	Human adenosine re
22	111.8	4.0	1316	15	AAQ56911	DNA encoding a gly
23	111.8	4.0	1316	18	AAT61679	Human alpha(1,3)-f
24	106.2	3.8	347	14	AAQ59752	Human brain Expres
25	103.6	3.7	795	17	AAT13798	Alpha-(1,3/1/4)-fu
26	102.4	3.6	2175	11	AAQ06691	CDX, a MILA involv
27	100.8	3.6	1256	21	AAF21134	Human low adenosin
28	100.8	3.6	1256	21	AAA35012	Human adenosine re
29	100.8	3.6	1400	13	AAQ31436	Encodes a HeLa cel
30	100.8	3.6	1488	15	AAQ56910	DNA encoding a gly
31	100.8	3.6	2175	12	AAQ14382	Clone 7.2 encoding
32	100.8	3.6	2175	24	AAS17082	Human cDNA clone 7
33	100.8	3.6	2861	12	AAQ14383	Clone 1 encoding 1
34	100.8	3.6	2861	21	AAF21133	Human low adenosin
35	100.8	3.6	2861	21	AAA35011	Human adenosine re
36	100.8	3.6	2861	24	AAS17083	Human cDNA clone 1
37	100.8	3.6	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
38	100.8	3.6	3647	15	AAQ56909	DNA encoding a gly
39	100.8	3.6	3647	18	AAT61678	Human alpha(1,3)-f
40	97.6	3.5	2134	18	AAT59506	Human myeloid deri
c 41	73	2.6	8033	22	AAS46308	Tumour suppressor
c 42	73	2.6	8033	24	AAS61117	Human gene regulat
c 43	69.2	2.5	6292	22	AAS46735	Tumour suppressor
c 44	67	2.4	5986	24	AAS61433	Human gene regulat
45	66.6	2.4	1578	23	ABL11867	Drosophila melanog

ALIGNMENTS

RESULT 1
AAZ92646
ID AAZ92646 standard; cDNA; 2822 BP.
XX
AC AAZ92646;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1289..2368
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

```
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 134-142; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

  Query Match             100.0%;  Score 2822;  DB 21;  Length 2822;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2822;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AAGCTTTGCTACTCAGAGAGTCTCATATGCTTTAAGAAAAACTACATAGCAAAAGAAGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGCTTTGCTACTCAGAGAGTCTCATATGCTTTAAGAAAAACTACATAGCAAAAGAAGGT 60

Qy      61 GCTTTTTTTAATGTTCACATATATTTAGGTAATTAAATGAAGGATAATTTCTCCCGTTTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTTTTTTTAATGTTCACATATATTTAGGTAATTAAATGAAGGATAATTTCTCCCGTTTT 120

Qy     121 TTGGAAGGAAACATCTAATGTTTCTATGCTAGTTTCTGTCAAAATTTCAAAATACTTTAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTGGAAGGAAACATCTAATGTTTCTATGCTAGTTTCTGTCAAAATTTCAAAATACTTTAA 180

Qy     181 TGAGTCTGAAATTTATAGGCCATATTATTTTCGTAAGAGAAGCAACATGAATTTCAAATG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGAGTCTGAAATTTATAGGCCATATTATTTTCGTAAGAGAAGCAACATGAATTTCAAATG 240

Qy     241 ACCCGTGAGAATTTTATCAATAAACAGTTTTTGCAAATATATTTTTTGTGTTCATAGCCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACCCGTGAGAATTTTATCAATAAACAGTTTTTGCAAATATATTTTTTGTGTTCATAGCCAT 300

Qy     301 AAAGAGGCATGAATGCCGAAGAGAAAAAAACTTGAAAACACTTTTTTCTTTAAAAAAAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAAGAGGCATGAATGCCGAAGAGAAAAAAACTTGAAAACACTTTTTTCTTTAAAAAAAAA 360

Qy     361 GCCTTCACTCATCTTTTTAGCCACAATCTGCTAAAATTCATGACATCAAAGTTGTTCTTC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCCTTCACTCATCTTTTTAGCCACAATCTGCTAAAATTCATGACATCAAAGTTGTTCTTC 420

Qy     421 ATGAAATATAGCCCTTTTTTTGAGAATTTGTTTATTTTTTATAATACCGATATCATCTTAAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGAAATATAGCCCTTTTTTTGAGAATTTGTTTATTTTTTATAATACCGATATCATCTTAAT 480

Qy     481 ACAAAGAAAAACTTTTTTAAAGCCCTTAGGCTATGTAAGTGAGAACGTTTTTCCAATGCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACAAAGAAAAACTTTTTTAAAGCCCTTAGGCTATGTAAGTGAGAACGTTTTTCCAATGCA 540

Qy     541 AATAGAACATTTGATTATTACATCTTGGTAGATATCGGATGTACTAACAAATTATTAAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AATAGAACATTTGATTATTACATCTTGGTAGATATCGGATGTACTAACAAATTATTAAAA 600

Qy     601 GTAAACAAAGAAAACTGTAAGAAAAATATATAAATAAAAACAAAACACTGTGAGTTTTGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTAAACAAAGAAAACTGTAAGAAAAATATATAAATAAAAACAAAACACTGTGAGTTTTGA 660

Qy     661 ATTTATAAAATTTGCCTGAGTGTACTTGTTTTTTAAAAATGTATCAAATCAAGAGTTGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ATTTATAAAATTTGCCTGAGTGTACTTGTTTTTTAAAAATGTATCAAATCAAGAGTTGCC 720

Qy     721 ATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTATAACG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTATAACG 780

Qy     781 AGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCATCTGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCATCTGA 840

Qy     841 ACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACTGGGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACTGGGA 900

Qy     901 GATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTAGCGA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTAGCGA 960

Qy     961 TGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACTGTTAAGGATAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACTGTTAAGGATAG 1020

Qy    1021 CATTGAAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTTTGGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CATTGAAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTTTGGG 1080

Qy    1081 AAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAATGGTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAATGGTT 1140

Qy    1141 TGATAACACTGAAAAGGAAAATGCTGTTGATTACTACTTGTTTATCTCATATTTATGATT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGATAACACTGAAAAGGAAAATGCTGTTGATTACTACTTGTTTATCTCATATTTATGATT 1200

Qy    1201 TTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCCCTTC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCCCTTC 1260

Qy    1261 TGTCTTTCTCTATTTCGTAGGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGTCTTTCTCTATTTCGTAGGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCC 1320

Qy    1321 ATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 ATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAA 1380

Qy    1381 ACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 ACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAA 1440

Qy    1441 AAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 AAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTG 1500

Qy    1501 GCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1501 GCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATG 1560

Qy   1561 CCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCG 1620

Qy   1621 AGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATG 1680

Qy   1681 GATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTT 1740

Qy   1741 GTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTT 1800

Qy   1801 GACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTG 1860

Qy   1861 GGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAA 1920

Qy   1921 AAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTT 1980

Qy   1981 GATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGA 2040

Qy   2041 TTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGG 2100

Qy   2101 ACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGA 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGA 2160

Qy   2161 TTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT 2220

Qy   2221 ATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGA 2280

Qy   2281 ATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 ATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGG 2340

Qy   2341 TAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATAT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATAT 2400

Qy   2401 TTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTG 2460

Qy   2461 TCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTA 2520

Qy   2521 AAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTA 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 AAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTA 2580

Qy   2581 TGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTT 2640

Qy   2641 TTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGT 2700

Qy   2701 TTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTAT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 TTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTAT 2760

Qy   2761 TACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAAT 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAAT 2820

Qy   2821 TC 2822
          ||
Db   2821 TC 2822
```

RESULT 2
AAZ92647

```
ID   AAZ92647 standard; cDNA; 2676 BP.
XX
AC   AAZ92647;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             295..1374
FT                   /*tag=  a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
```

```
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match             54.6%;  Score 1542;  DB 21;  Length 2676;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-293;
  Matches 1542;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 346

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 406

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 466

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 526

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 586

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    587 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 646

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    647 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 706

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    707 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 766

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    767 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 826

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    827 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 886

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    887 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 946

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    947 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 1006

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1007 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1066

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1067 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1126

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1127 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1186

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1246

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1247 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1306

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1366

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1367 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 1426

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1427 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 1486

Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1487 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 1546

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1547 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 1606

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1607 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 1666

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1667 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 1726

Qy   2721 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 2780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1727 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 1786

Qy   2781 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 2822
           |||||||||||||||||||||||||||||||||||||||||
Db   1787 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 1828


RESULT 3
AAZ92672
ID   AAZ92672 standard; cDNA; 2036 BP.
XX
AC   AAZ92672;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
```

XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Example 8; Page 162-164; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC (AAY80995), and sequences AAZ92670-Z92673 are murine
CC alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC amplification of cDNA ends).
XX
SQ Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

Query Match 38.7%; Score 1092.2; DB 21; Length 2036;
Best Local Similarity 85.8%; Pred. No. 2.6e-205;
Matches 1309; Conservative 0; Mismatches 198; Indels 18; Gaps 8;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          ||||||||||||||||||||||||||||| ||||||||||||||| |||| |||||||| |
Db      1 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 60

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db     61 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 120

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||| || || || |||||||||||||||||||| ||||| ||||| ||||
Db    121 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 180

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          |||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||||
Db    181 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 240

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          |||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |
Db    241 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 300

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640
          || |||||||||||||||||| || ||||| || ||  |||||||||| ||||||||||
Db    301 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 360

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700
          | || |||||||||| || |||||||||| |||||||||||||||||||||||| || |
Db    361 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 420

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760
          |||| |||||||| |||||||||||||||||||| |||||||| |||||||||| ||||
Db    421 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 480

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820
          | |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    481 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 540

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880
          | |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Db    541 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 600

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940
          |||||||||||| |||||||||||| ||||| ||||| || |||||||||| ||||
Db    601 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 660

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000
          | || |||||||| |||||||| || || |||||||| |||||| |||||||| |||||
Db    661 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 720

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060
          |||||||||||||||| |||||||| ||||| |||| |||||||||| |||||||| |
Db    721 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 780

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120
          ||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||||||||
Db    781 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 840

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||
Db    841 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 900

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240
          ||||||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db    901 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 960

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300
          |||| |||||||| |||||||| || ||||| ||||||||||||||||||||  |||| |
Db    961 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1020

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360
          |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1080

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420
          |||||| ||| | | ||||| |  ||||    |  ||| |  || |||||| || || |||
Db    1081 GGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAA 1136

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480
          || || ||||||||||||      ||||  || | |||| | |||  |||| ||||    |
Db    1137 GTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGC 1191

Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540
          |||| || | | | ||||||||||||||| | |||||||| || |||| || ||||||| 
Db    1192 CCTCCCTTGAGGACCATGTATATTTTGGTGAA-ATTTTTAAGAGATCAGAATTAGCAATC 1250

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600
          | || ||| |||||||||||||||||||||  | || | | |||||||| | ||||||
Db    1251 ACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTT 1309

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660
          |||  ||| | |||||| |||||||| |||| ||   |||| |||||| |  ||| |||
Db    1310 ATTTGTCACTCTCATTTATAAACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTA 1366

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720
          |  || || ||   |  ||||||| |||   ||||| | |||||||||||||||| || |
```

```
Db   1367 AATTTATGATT---TACTTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAA 1423

Qy   2721 AGATGCACATCTTAAAGTATGA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAAC 2779

Db   1424 AAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAG 1483

Qy   2780 TTTGCATACTATAACAGAGGAAGAA 2804

Db   1484 CTTGTATAATGCTAACAAAGAAGAA 1508


RESULT 4
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag=  a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match             38.7%;  Score 1092.2;  DB 21;  Length 2170;
  Best Local Similarity   85.8%;  Pred. No. 2.7e-205;
  Matches 1309;  Conservative    0;  Mismatches  198;  Indels   18;  Gaps    8;

Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340

Db    107 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 166

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400

Db    167 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 226

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460

Db    227 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 286

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520

Db    287 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 346

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580

Db    347 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 406

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640

Db    407 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 466

Qy   1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700

Db    467 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 526

Qy   1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760

Db    527 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 586

Qy   1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820

Db    587 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 646

Qy   1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880

Db    647 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 706

Qy   1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940

Db    707 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 766

Qy   1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000

Db    767 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 826

Qy   2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060

Db    827 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 886

Qy   2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120

Db    887 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 946

Qy   2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180
```

```
Db    947 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 1006

Qy   2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240

Db   1007 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 1066

Qy   2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300

Db   1067 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1126

Qy   2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360

Db   1127 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1186

Qy   2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420

Db   1187 GGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAA 1242

Qy   2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480

Db   1243 GTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGC 1297

Qy   2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540

Db   1298 CCTCCCTTGAGGACCATGTATATTTTGGTGAA-ATTTTTAAGAGATCAGAATTAGCAATC 1356

Qy   2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600

Db   1357 ACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTT 1415

Qy   2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660

Db   1416 ATTTGTCACTCTCATTTATAAACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTA 1472

Qy   2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720

Db   1473 AATTTATGATT---TACTTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAA 1529

Qy   2721 AGATGCACATCTTAAAGTATGA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAAC 2779

Db   1530 AAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAG 1589

Qy   2780 TTTGCATACTATAACAGAGGAAGAA 2804

Db   1590 CTTGTATAATGCTAACAAAGAAGAA 1614


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag=  a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             4.5%;  Score 127.4;  DB 18;  Length 1814;
  Best Local Similarity   52.1%;  Pred. No. 8e-16;
  Matches  334;  Conservative    0;  Mismatches  301;  Indels    6;  Gaps    2;

Qy   1557 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 1616

Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy   1617 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 1676

Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy   1677 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 1736

Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy   1737 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 1796

Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy   1797 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 1856

Db    788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy   1857 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 1916

Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy   1917 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 1976

Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy   1977 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 2036

Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy   2037 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 2093

Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy   2094 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 2153
```

```
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy   2154 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 2194
          |||| || |  | |||| || |||| || ||    |  |||
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag= a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag= b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag= c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag= d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag= e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag= f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag= g
FT                   /number= 3
FT   exon            2067..3552
FT                   /*tag= f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             4.5%;  Score 127;  DB 18;  Length 3594;
  Best Local Similarity   52.3%;  Pred. No. 1.1e-15;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy   1557 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 1616
           | |||| ||| |   |  ||||   |||    ||   |  |||| ||  |  |||| |
Db   2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336

Qy   1617 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 1676
           |||| ||  |          |   |     |||| | | |    |||||||  |   |||
Db   2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy   1677 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 1736
             ||| | |||       |||||||| || | | | || |          | || |   |
Db   2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy   1737 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 1796
           | | || |||  |   ||||  || || ||||||||||||||||   || || || ||
Db   2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy   1797 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 1856
           ||||  |     | |   ||   |    | | ||   ||||   |   |||
Db   2517 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy   1857 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 1916
           ||||||  | || || |   |     |||    | |  |||   |||    ||||
Db   2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy   1917 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 1976
           |     |  |  |  ||   | ||| |  ||    |||   .   |    | |||
Db   2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy   1977 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 2036
           | || | ||  |  |   |     || || ||  |||||| || |||   ||
Db   2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy   2037 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 2093
           ||| |||||||| || ||| | |    ||||||  |   ||||||  |||||||  || |
Db   2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy   2094 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 2153
             ||||||||  || |||  | ||| |||  || | | ||| ||||| | ||  | || |
Db   2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy   2154 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
          |||| || |  | |||| || |||| || ||
Db   2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904


RESULT 7
AAQ77732
```

```
ID   AAQ77732 standard; cDNA to mRNA; 1701 BP.
XX
AC   AAQ77732;
XX
DT   22-JUN-1995  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; sialylated Lewis blood group antigen;
KW   Lewis x; Lewis a; glycosylation; glycoprotein; glycolipid; ds.
XX
OS   Homo sapiens (THP-1 cells).
XX
FH   Key             Location/Qualifiers
FT   CDS             123..1151
FT                   /*tag= a
FT                   /product= alpha-1,3-fucosyltransferase
XX
PN   WO9423021-A.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994;   94WO-JP00496.
XX
PR   29-MAR-1993;   93JP-0069016.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hanai N,  Kurata K,  Nishi T,  Sasaki K;
XX
DR   WPI; 1994-333183/41.
DR   P-PSDB; AAR63215.
XX
PT   Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
PT   of active sugar chains, e.g. sialylated Lewis X
XX
PS   Claim 3; Page 99-102; 155pp; Japanese.
XX
CC   The sequence coding for alpha-1,3-fucosyltransferase was isolated
CC   from a human THP-1 cell cDNA library. The enzyme catalyses the
CC   addition of fucose via a 1,3-linkage to N-acetylglucosamine
CC   contained in glycoproteins and glycolipids, e.g. for sialylation of
CC   Lewis blood group antigens.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             4.1%;  Score 114.4;  DB 15;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 2.8e-13;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy   1544 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 1603
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy   1604 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 1663
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy   1664 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 1723
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy   1724 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 1783
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy   1784 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 1843
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy   1844 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 1903
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy   1904 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 1963
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy   1964 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 2023
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy   2024 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 2080
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy   2081 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 2140
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy   2141 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973


RESULT 8
AAF21136
ID   AAF21136 standard; DNA; 1701 BP.
XX
AC   AAF21136;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2703.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1034; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
```

```
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1701 BP; 290 A; 527 C; 589 G; 295 T; 0 other;

  Query Match             4.1%;  Score 114.4;  DB 21;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 2.8e-13;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy    1544 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 1603
           | |  ||||    | ||||| || |   |  ||||   |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    1604 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 1663
           ||  |  |||| |||||  ||  |          | | |  |  | || | |       ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    1664 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 1723
           || |    |||   ||| | |||       ||||  || || |  ||||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    1724 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 1783
           ||   | |      | | || |||   |   ||||  ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    1784 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 1843
           |||||  ||||||  | ||  |     |       ||||    |       |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    1844 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 1903
              |  |||    ||||||  || || ||||   |      |||     | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    1904 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 1963
           |||    ||||          |     |  ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    1964 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 2023
           |    |  |      || | || ||| |  |     |     || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    2024 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 2080
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    2081 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 2140
            |||| || || || ||||| ||  |  |||  | |||||||    |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    2141 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

```
RESULT 9
AAA35014
ID   AAA35014 standard; DNA; 1701 BP.
XX
AC   AAA35014;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2703.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 957-958; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             4.1%;  Score 114.4;  DB 21;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 2.8e-13;
```

```
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy   1544 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 1603
          | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy   1604 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 1663
          ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy   1664 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 1723
          || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy   1724 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 1783
          ||   | |       | | || |||   |   ||||  ||||| || || || || |||
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy   1784 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 1843
          |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy   1844 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 1903
             |  |||    |||||| || || ||||   |      |||     | | || |
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy   1904 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 1963
          |||    ||||            |     |  ||  ||  || |  ||   ||||
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy   1964 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 2023
          |    |  |       || | || ||| |  |    |      || || || ||||||||
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy   2024 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 2080
          || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy   2081 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 2140
          |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy   2141 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
          | ||| | |||||||| || | |  |||  || | ||||| ||
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973


RESULT 10
AAF21137
ID   AAF21137 standard; DNA; 6944 BP.
XX
AC   AAF21137;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2704.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1034-1036; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

  Query Match             4.1%;  Score 114.4;  DB 21;  Length 6944;
  Best Local Similarity   50.8%;  Pred. No. 3.7e-13;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy   1544 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 1603
          | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db   5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy   1604 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 1663
          ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db   5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy   1664 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 1723
          || |     |||   ||| | |||       |||| || || |  ||||| | |
Db   5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy   1724 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 1783
          ||   | |       | | || |||   |   ||||  ||||| || || || |||
Db   5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy   1784 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 1843
          |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db   5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872
```

```
Qy   1844 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 1903
Db   5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy   1904 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 1963
Db   5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy   1964 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 2023
Db   5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy   2024 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 2080
Db   6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy   2081 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 2140
Db   6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy   2141 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
Db   6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

RESULT 11
AAA35015
ID AAA35015 standard; DNA; 6944 BP.
XX
AC AAA35015;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2704.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 958-960; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

Query Match 4.1%; Score 114.4; DB 21; Length 6944;
Best Local Similarity 50.8%; Pred. No. 3.7e-13;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy   1544 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 1603
Db   5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy   1604 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 1663
Db   5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy   1664 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 1723
Db   5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy   1724 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 1783
Db   5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy   1784 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 1843
Db   5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy   1844 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 1903
Db   5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy   1904 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 1963
Db   5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy   1964 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 2023
Db   5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy   2024 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 2080
Db   6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy   2081 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 2140
Db   6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy   2141 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
Db   6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

RESULT 12

```
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             4.0%;  Score 113.4;  DB 15;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 4.4e-13;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
           || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
           |      ||      |  ||| |  ||||  | |||||      ||||    |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
           |||     ||| || ||||           | | | |      |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
           |  |  |   |||      |  |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
           |||||  ||     ||  ||| |     ||             |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
           |       |      |      |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
           |||||  |||| || || ||  | |||   || |||||||| || |||||  |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
           ||  |   |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
           |    | ||| | ||  | ||||| || ||||| || |    |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
           |||  |||  ||||  | |||||  || |       |||||  ||  ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   2246 AAG 2248
           ||
Db   1095 GAG 1097


RESULT 13
AAT61680
ID   AAT61680 standard; DNA; 1654 BP.
XX
AC   AAT61680;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..1208
FT                   /*tag=  a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13643.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
```

PS Example 6; Page 291-292; 329pp; English.
XX
CC A DNA clone (AAT61680) codes for an enzyme (AAW13643) that functions as
CC a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC precursors but not type I glycoconjugates. It was identified in a
CC human lambda phage genomic DNA library using a probe obtd. from
CC Lewis enzyme cDNA (see also AAT61675). The Fuc-TVI DNA can be used
CC to construct animal cell lines with specific capabilities with
CC respect to post-translational modification of the oligosaccharides
CC of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC use in oligosaccharide prodn., to genotype individuals at this
CC fucosyltransferase locus, and in the prodn. of chimeric
CC fucosyltransferases (see also AAW14512-32).
XX
SQ Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

Query Match 4.0%; Score 113.4; DB 18; Length 1654;
Best Local Similarity 51.2%; Pred. No. 4.4e-13;
Matches 401; Conservative 0; Mismatches 361; Indels 21; Gaps 5;

```
Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
          || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
            |     ||    |  ||| |  |||| | |||||       ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
          || ||  |  | || ||||||||  |||  |  | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
          | |||||  |     |||  |||||| ||| | |   |||| || |||| |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
          ||  ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
          |||    ||| || ||||             | | | |      |  | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
          |  |  |   |||     |  |||||  |||||  |||    ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
          |||||  ||    ||  ||| |   ||              |  |  |  |    |||||  |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
            |     |     |    |   ||   ||||    ||  |  ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
          |||||  |||| || || ||  | |||    || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
          ||  |   |||  | || || || ||||| ||    ||      |||||  ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
          |  ||| | ||  | ||||| || ||||| || |    |     ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
          |||  ||| ||||  | |||||  || |       |||||  || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094
```

Qy 2246 AAG 2248
||
Db 1095 GAG 1097

RESULT 14
AAQ56912
ID AAQ56912 standard; DNA; 1086 BP.
XX
AC AAQ56912;
XX
DT 26-JUL-1994 (first entry)
XX
DE Lewis blood group fucosyltransferase DNA.
XX
KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW surface; oligosaccharide; ss.
XX
PN WO9402616-A.
XX
PD 03-FEB-1994.
XX
PF 20-JUL-1993; 93WO-US06703.
XX
PR 20-JUL-1992; 92US-0914281.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Lowe JB;
XX
DR WPI; 1994-048874/06.
XX
PT DNA fragment encoding a glycosyltransferase - can be used for in
PT vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT blood gp. determinants, to protect against transplant rejection
XX
PS Disclosure; Fig 6; 249pp; English.
XX
CC The sequence is that encoding a Lewis blood group fucosyl
CC transferase (Fuc-TIII). The enzyme produced by the DNA may be
CC non glycosylated. This prevents premature loss of enzyme activity.
CC It can also be used in in vitro reactions to modify cell surface
CC oligosaccharide mols. e.g. blood group determinants.
CC See also AAQ56905-11.
XX
SQ Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

Query Match 4.0%; Score 112.2; DB 15; Length 1086;
Best Local Similarity 51.6%; Pred. No. 7e-13;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | |||||       |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
            | || ||||  |   |   |  | ||    | || |              |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
            ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
            | || || ||  ||  | | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |  ||  |   |  |   |                  | |  |  |  |  |  |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
            |  |||||  |||||  |||   |||||||| || ||  .  |||||  ||     || |||
```

```
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
          |    ||           |  |  |  |    |||||  |   |               |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972
```

RESULT 15
AAT80111
ID AAT80111 standard; cDNA; 2042 BP.
XX
AC AAT80111;
XX
DT 01-DEC-1997 (first entry)
XX
DE Fucosyltransferase III coding sequence.
XX
KW P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW adult respiratory distress syndrome; tumour metastasis;
KW fucosyltransferase III; acetylglucosaminyltransferase;
KW autoimmune disorder; inflammatory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO9706176-A2.
XX
PD 20-FEB-1997.
XX
PF 02-AUG-1996; 96WO-US12820.
XX
PR 17-MAY-1996; 96US-0649802.
PR 03-AUG-1995; 95US-0510920.
PR 15-MAY-1996; 96US-0017794.
XX
PA (OKLA) UNIV OKLAHOMA STATE.
XX
PI Cummings RD, Mcever RP, Moore KL;
XX
DR WPI; 1997-154206/14.
XX
PT New O-glycan(s) derived from P-selectin glyco:protein ligand
PT (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT for treating inflammatory or autoimmune disorders or tumours
XX
PS Disclosure; Page 80-81; 96pp; English.
XX
CC This sequence represents the coding sequence for the human
CC fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC protein expressed on the surfaces of activated platelets and endothelium
CC in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC high affinity P-selectin ligand produced by leukocytes. Binding of
CC P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC highly extended extracellular domain, which is a feature of mucin-like
CC proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC O-glycans, and O-glycanated products of the invention. The products are
CC used for inhibiting the binding of P-selectin and other selectins to
CC PSGL-1. They can be used for the treatment of inflammation, injury
CC resulting from ischaemia and reperfusion, bacterial sepsis and
CC disseminated intravascular coagulation, adult respiratory distress
CC syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC other autoimmune or inflammatory disorders. The products can also be used
CC in the production of antibodies and in diagnostic applications.
XX
SQ Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

Query Match 4.0%; Score 112.2; DB 18; Length 2042;
Best Local Similarity 51.6%; Pred. No. 7.9e-13;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | ||||||     |||||    |  |  |  | ||  | | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
          | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
             ||||| ||| | || ||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |   ||  |   | |   |          |  | |  |  |  |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
          |    ||           |  |  |  |    |||||  |   |               |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |       ||||| || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 5, 2002, 23:12:59
Job time : 476.687 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 2362.78 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-4
Perfect score: 2822
Sequence: 1 aagctttgctactcagagag.....aacatgttgcgattgaattc 2822

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	603.2	21.4	716	9	AV725071	AV725071 AV725071
	2	575.6	20.4	697	9	BB397507	BB397507 BB397507
c	3	573.2	20.3	579	10	BM021081	BM021081 ie75f03.x
c	4	557.8	19.8	561	12	AQ321598	AQ321598 RPCI11-93
	5	492.8	17.5	607	10	BI102760	BI102760 602888338
c	6	448	15.9	501	12	B41073	B41073 HS-1052-B2-
	7	430	15.2	649	9	AV338075	AV338075 AV338075
c	8	409.2	14.5	548	12	AQ772584	AQ772584 HS_5568_A
c	9	296.8	10.5	1101	12	CNS059ST	AL327638 Tetraodon
	10	292.6	10.4	648	9	BB595164	BB595164 BB595164
	11	280.4	9.9	625	9	BB649967	BB649967 BB649967
	12	269.6	9.6	767	9	BB640911	BB640911 BB640911
	13	251	8.9	568	10	BM021317	BM021317 ie75f03.y
	14	227.6	8.1	628	9	BB651409	BB651409 BB651409
c	15	184.2	6.5	570	10	BM082484	BM082484 fu25e08.x
c	16	177	6.3	1100	12	CNS0579Z	AL324368 Tetraodon
c	17	170.2	6.0	1016	12	CNS0429P	AL271222 Tetraodon
	18	165	5.8	844	12	CNS01V8V	AL168808 Tetraodon
c	19	157.4	5.6	887	12	CNS02Q0S	AL208693 Tetraodon
	20	155.4	5.5	651	10	BJ036855	BJ036855 BJ036855
c	21	151.6	5.4	666	10	BJ050300	BJ050300 BJ050300
c	22	151.2	5.4	711	12	CNS034XF	AL228012 Tetraodon
	23	146.6	5.2	640	10	BJ073227	BJ073227 BJ073227
	24	146	5.2	533	10	BI604066	BI604066 603243886
	25	144.6	5.1	670	9	BB626126	BB626126 BB626126
	26	138	4.9	199	9	BB398103	BB398103 BB398103
c	27	125.2	4.4	674	10	BE605503	BE605503 fl10h04.x
	28	120.6	4.3	971	12	CNS047IV	AL278032 Tetraodon
c	29	120.4	4.3	615	10	BI326811	BI326811 STR00078
	30	119.8	4.2	284	9	BB270301	BB270301 BB270301
c	31	117.8	4.2	1023	12	CNS02ANG	AL188773 Tetraodon
	32	115.6	4.1	317	9	BB272545	BB272545 BB272545
	33	106.2	3.8	347	10	M78549	M78549 EST00697 Fe
c	34	103	3.6	702	10	BJ043578	BJ043578 BJ043578
	35	102.8	3.6	487	10	BE863396	BE863396 UI-M-BH0-
c	36	96.2	3.4	1101	12	CNS0039G	AL063921 Drosophil
	37	93.8	3.3	863	12	CNS02ROF	AL210840 Tetraodon
	38	90.4	3.2	720	10	BF180380	BF180380 601804564
	39	90	3.2	551	10	BJ035701	BJ035701 BJ035701
	40	89.4	3.2	650	9	AV375511	AV375511 AV375511
	41	89	3.2	1101	12	CNS00EVL	AL069706 Drosophil
	42	87.4	3.1	601	10	BJ036679	BJ036679 BJ036679
	43	87.4	3.1	605	10	BJ028804	BJ028804 BJ028804
c	44	87	3.1	1056	12	CNS01VL7	AL169252 Tetraodon
	45	86.8	3.1	304	9	AU081478	AU081478 AU081478

ALIGNMENTS

RESULT 1
AV725071
LOCUS AV725071 716 bp mRNA linear EST 16-OCT-2000
DEFINITION AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION AV725071
VERSION AV725071.1 GI:10830091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 716)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA HTC clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .716
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBLB02"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 219 a 113 c 133 g 249 t 2 others
ORIGIN

```
  Query Match           21.4%;  Score 603.2;  DB 9;  Length 716;
  Best Local Similarity  99.2%;  Pred. No. 2.8e-78;
  Matches  605;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy   1947 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 2006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy   2007 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 2066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy   2067 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 2126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy   2127 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 2186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy   2187 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 2246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy   2247 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 2306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy   2307 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 2366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy   2367 AAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTG 2426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTG 480

Qy   2427 AGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTC 2486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTC 540

Qy   2487 TAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCA 2546
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    541 TAGGGTAACGTGTATATCTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCA 600

Qy   2547 TTCGGTTTTA 2556
          ||    ||||
Db    601 TTNNCGTTTA 610


RESULT 2
BB397507
LOCUS       BB397507                 697 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA
            clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for
            alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   BB397507
VERSION     BB397507.2  GI:16411808
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 15, 2000 this sequence version replaced gi:9216903.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN

  Query Match           20.4%;  Score 575.6;  DB 9;  Length 697;
  Best Local Similarity  91.2%;  Pred. No. 2.8e-74;
  Matches  622;  Conservative    0;  Mismatches   59;  Indels    1;  Gaps    1;

Qy   1692 ATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGA 1751
          | || || ||||| |||||||| ||| ||||||||||| ||| |||||||| |||||||
Db      4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy   1752 CTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCA 1811
```

```
Db     63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy   1812 CAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTA 1871

Db    123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy   1872 ACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAA 1931

Db    183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy   1932 TCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCA 1991

Db    243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy   1992 TATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGG 2051

Db    303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy   2052 AAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGG 2111

Db    363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy   2112 AAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTC 2171

Db    423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy   2172 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTT 2231

Db    483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy   2232 ACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCAT 2291

Db    543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy   2292 GTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 2351

Db    603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy   2352 AATGGTTTTGGAATTAAAATTT 2373

Db    663 AATGGTTTTGGAATTAAAGTGT 684
```

```
RESULT 3
BM021081/c
LOCUS       BM021081                579 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021081
VERSION     BM021081.1  GI:16535437
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
BASE COUNT      203 a    103 c     93 g    180 t
ORIGIN
```

```
  Query Match           20.3%;  Score 573.2;  DB 10;  Length 579;
  Best Local Similarity  99.5%;  Pred. No. 6.6e-74;
  Matches  575;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   2060 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 2119

Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy   2120 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 2179

Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy   2180 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 2239

Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy   2240 TGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCT 2299

Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy   2300 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 2359

Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy   2360 TGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCA 2419

Db    279 TGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCA 220

Qy   2420 AGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCA 2479

Db    219 AGTATTGAAAATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCA 160

Qy   2480 CCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAAT 2539

Db    159 CCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAAT 100

Qy   2540 CATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATT 2599

Db     99 CATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATT 40

Qy   2600 TATTCTTCATTATCATTTGTAAACATTGCTTTTTCACA 2637
```

```
          ||||||||||||||||||||||||||||||||||||
Db      39 TATTCTTCATTATCATTTGTAAACATTGCTTTTTCACA 2


RESULT 4
AQ321598/c
LOCUS       AQ321598                561 bp    DNA     linear   GSS 06-MAY-1999
DEFINITION  RPCI11-93E9.TV RPCI-11 Homo sapiens genomic clone RPCI-11-93E9, DNA
            sequence.
ACCESSION   AQ321598
VERSION     AQ321598.1  GI:4054332
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-93E9.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Homo sapiens"
                     /db_xref="GDB:7535432"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-93E9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      200 a     92 c     89 g    180 t
ORIGIN

  Query Match           19.8%;  Score 557.8;  DB 12;  Length 561;
  Best Local Similarity   99.6%;  Pred. No. 1.1e-71;
  Matches  559;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     717 TGCCATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTAT 776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 TGCCATTTTTTTACCAGATGATTATACTAATGCATTTGCAAACTGTGAAGGTAAATTTAT 502

Qy     777 AACGAGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCAT 836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 AACGAGAAAAACAAAAACATAACGTTTTTAAAAACAGTTTAGTTTTCTAACGAATTCCAT 442

Qy     837 CTGAACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACT 896
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 CTGAACCACAAAGTTATGGTACAGTCTCAATATGCGAGTAAATAAGTGATTATATCAACT 382

Qy     897 GGGAGATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTA 956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 GGGAGATGTTCACAAATCTGATGTGGTTACAAAATACTAAAAATTTTAGTCATGGCTTTA 322

Qy     957 GCGATGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACTGTTAAGG 1016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||

Db     321 GCGATGACAGTCTAGGTTGTGCCACTTTTCTACAAGCAAATAAAAGTAGAACAGTTAAGG 262

Qy    1017 ATAGCATTGAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTT 1076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ATAGCATTGAAAAAAACGTTACTTGGTTTGAAGATATGCTGGTACTTTTAGAAGAGTTTT 202

Qy    1077 TGGGAAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAAT 1136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 TGGGAAACTTCCTTTTGAAAGCTGTTTCTTGGCTTTGAGGAAAGTTCTGTTTCTCTCAAT 142

Qy    1137 GGTTTGATAACACTGAAAAGGAAAATGCTGTTGATTACTACTTGTTTATCTCATATTTAT 1196
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     141 GGTTTGATAACACTGAAAAGGAAAATGCAGTTGATTACTACTTGTTTATCTCATATTTAT 82

Qy    1197 GATTTTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCC 1256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GATTTTAAATATATCTCCAATATATTTATGATTTTCTCTTCATTCCCACCGCTACCTCCC 22

Qy    1257 CTTCTGTCTTTCTCTATTTCG 1277
           |||||||||||||||||||||
Db      21 CTTCTGTCTTTCTCTATTTCG 1


RESULT 5
BI102760
LOCUS       BI102760                607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: l  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN

  Query Match           17.5%;  Score 492.8;  DB 10;  Length 607;
  Best Local Similarity   90.1%;  Pred. No. 2.9e-62;
  Matches  549;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy    1511 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACA 1570
           |||||||||||| |||||||||||||||||||||||| |||||||| ||||||||||||
Db       1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60
```

```
Qy   1571 ACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGT 1630
          || ||||| ||| |||||||||||||||||||| || ||||| || ||  ||||||||||
Db     61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy   1631 TGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATG 1690
          |||||||||| || |||||||||| || |||||||||| |||||||||| |||||||
Db    120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy   1691 AATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTG 1750
          ||||| || ||||| |||||||| |||||||||||||||||| |||||||| ||||||
Db    179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy   1751 ACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGC 1810
          ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||| |||
Db    239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy   1811 ACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGT 1870
          |||||||||||| |||||||||||||||||| ||||| ||||||||||||||||||||
Db    299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy   1871 AACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAA 1930
          |||||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||
Db    359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy   1931 ATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACC 1990
          ||||| ||||| || |||||||| |||||||. || || ||||||||| ||||||| |||
Db    419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy   1991 ATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACG 2050
          |||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||
Db    479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy   2051 GAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGG 2110
          |||||||| |||||||| ||| ||||||| || |||||  ||| ||||| |||||||||
Db    539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy   2111 GAAAACTAT 2119
          |||||||||
Db    599 GAAAACTAT 607
```

```
RESULT 6
B41073/c
LOCUS       B41073                 501 bp    DNA     linear   GSS 18-OCT-1997
DEFINITION  HS-1052-B2-G01-MF.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 774 Col=2 Row=N, DNA sequence.
ACCESSION   B41073
VERSION     B41073.1  GI:2545325
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 774  row: N  column: 2
            Class: BAC ends
            High quality sequence stop: 501.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=CT 774 Col=2 Row=N"
                     /clone_lib="CIT Human Genomic Sperm Library C"
                     /sex="M"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      179 a     95 c     83 g    143 t      1 others
ORIGIN
```

```
  Query Match           15.9%;  Score 448;  DB 12;  Length 501;
  Best Local Similarity   99.8%;  Pred. No. 9.4e-56;
  Matches  448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   2218 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 2277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 442

Qy   2278 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 2337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 382

Qy   2338 TGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAA 2397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 TGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAA 322

Qy   2398 TATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTT 2457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 TATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTT 262

Qy   2458 GTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTT 2517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 GTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTT 202

Qy   2518 TTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAA 2577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAA 142

Qy   2578 TTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACA 2637
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    141 TTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTNCACA 82

Qy   2638 TTTTTGTAGTTGTCCATAATGTAAGCTTG 2666
          |||||||||||||||||||||||||||||
Db     81 TTTTTGTAGTTGTCCATAATGTAAGCTTG 53
```

```
RESULT 7
AV338075
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
```

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430407H24"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory bulb"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      209 a    105 c    117 g    218 t
ORIGIN
```

```
  Query Match             15.2%;  Score 430;  DB 9;  Length 649;
  Best Local Similarity   83.8%;  Pred. No. 3.5e-53;
  Matches  548;  Conservative    0;  Mismatches   95;  Indels   11;  Gaps    5;

Qy   1993 ATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGA 2052
          ||| |||||||| |||  |||||| |||||||| ||||| |||||||||||||||||| ||
Db      2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61

Qy   2053 AAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGA 2112
          |||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy   2113 AAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCC 2172
          ||||||||||||||||||||||| |||||||||||||||||||||| |||||||||
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy   2173 CAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTA 2232
          ||||||| |||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy   2233 CTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATG 2292
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy   2293 TTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 2352
            |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy   2353 ATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATA 2412
          |||||||||||||| ||| | | ||||| |  ||||   |  ||| |  || |||||| |
Db    362 ATGGTTTTGGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACA 417

Qy   2413 TCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATT 2472
          | || ||||| || ||||||||||||     ||||  || | |||| | |||  |||| |
Db    418 TAATTCAAGTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAAT 472

Qy   2473 TTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCAT 2532
          |||   ||||| || | | | | ||||||||||||| || |||||||| || |||| ||
Db    473 TTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGT-GAAATTTTTAAGAGATCAGAAT 531

Qy   2533 GAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAG 2592
          |||||||| || ||| |||||||||||||||||||||  | || | | ||||||||
Db    532 TAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA- 590

Qy   2593 AGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAG 2646
          | ||||||||  ||| | |||||| |||||||| |||| ||  || | || |
Db    591 AATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAGTTG 644
```

```
RESULT 8
AQ772584/c
LOCUS       AQ772584                548 bp    DNA     linear   GSS 29-JUL-1999
DEFINITION  HS_5568_A2_D09_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1144 Col=18 Row=G, DNA sequence.
ACCESSION   AQ772584
VERSION     AQ772584.1  GI:5652312
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
```

```
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1144  row: G  column: 18
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 548.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1144 Col=18 Row=G"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      190 a      94 c      89 g     168 t       7 others
ORIGIN

  Query Match             14.5%;  Score 409.2;  DB 12;  Length 548;
  Best Local Similarity   93.2%;  Pred. No. 3.8e-50;
  Matches  426;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy    2366 TAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATT 2425
           |||| |||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db     547 TAAATTTTTTCATCACTGGCACACTTGATAAATATTTTGATCAGATATCATCCAAGTATT 488

Qy    2426 GAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCT 2485
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     487 GAGGATAAGAAGAGATGCAACATCCTACTTTTGTGTCACAATTTATTTTTATCACCCTCT 428

Qy    2486 CTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCC 2545
           |||||||||||||||||||||||||| |||||| |||||||||| | ||||||||||||
Db     427 CTAGGGTAACGTGTATATTTTGGTGGAAATTTTTTAAAGCTCAGCTTAAGCAATCATTCC 368

Qy    2546 ATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCT 2605
           |||||||||||||| |||| |||||||| ||||||||||||||||||||||||||||| |
Db     367 ATTCGGTTTTAAATGATCCGGTATATACGCAATTATGTGCACTGGAGAGTAATTTATTTT 308

Qy    2606 TCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTT 2665
           |||||||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||
Db     307 TCATTATCATATGTAAGCATAGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTT 248

Qy    2666 GTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATG 2725
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||
Db     247 GTGGTTTGATACTTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAAATG 188

Qy    2726 CACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCA 2785
           |||||||||||||||| | ||||| | |||||||||||||| ||||||||||||||||||
Db     187 CACATCTTAAAGTATGTAGAATTTAAAATAAGTATTACAATGTGTAGTTCCAACTTTGCA 128

Qy    2786 TACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 2822
           |||||| ||| || |||||||||||| ||| |||| |
Db     127 TACTATGACANAGNAAGAACATGTTGNGATGGAATGC 91


RESULT 9
CNS059ST/c
LOCUS       CNS059ST                1101 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            045P22 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL327638
VERSION     AL327638.1  GI:8221227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="045P22"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA045DH11A1~end : T3"
BASE COUNT      287 a     249 c     269 g     278 t      18 others
ORIGIN

  Query Match             10.5%;  Score 296.8;  DB 12;  Length 1101;
  Best Local Similarity   58.3%;  Pred. No. 5.8e-34;
  Matches  603;  Conservative    0;  Mismatches  423;  Indels    9;  Gaps    5;

Qy    1338 TTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGA 1397
           | ||| |||| || |   || |         | || |||   || ||  ||| ||||| |
Db    1064 TCATCGTGGGGTGCTNTGTGACTGTGTNNTTTATGTACTNTAATCCATCCACCAGCT-GN 1006

Qy    1398 TCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCA 1457
           | | | ||||  ||||  |||          |  | |  | || ||     | || ||||
Db    1005 TTTCCGGTCCTGTGGAGGCAGGACTGAGCTCGGAGCATGTCAAGAATCAACTATCAACCA 946

Qy    1458 AAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCT 1517
           | | |||| |    ||    || | ||| ||| ||||| | ||||| || || || ||||
Db     945 AGAGTGATAAA---AACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCT 889

Qy    1518 TTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACC 1577
           | ||||| |     ||| | ||    |||  | |  | || ||  |  | ||  |||||
Db     888 TCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACA 829

Qy    1578 GTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATC 1637
           |    || |  ||||| ||  |||  ||| |  |||| ||| |||||||||    | || |
Db     828 GAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACC 769

Qy    1638 TGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGG 1697
           || | ||  | ||        |   | ||  || ||||||||||||| ||| |  |  |||
Db     768 TGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGG 709

Qy    1698 AATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGA 1757
           | || ||  | ||| |  |      |   ||| || |||  | |||| |||||||| ||||
Db     708 AGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGA 649

Qy    1758 CTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATC 1817
           || || |  ||||| || ||         ||  |   || |     ||   |||  ||
Db     648 ACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATG 589

Qy    1818 CCTTCGTG---TTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 1874
           |    | |   ||||      || |||||| | ||| |||| || ||| |||| || ||||
Db     588 CTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACT 529
```

```
Qy   1875 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 1934
          ||||||  || |||||  |||| || |  ||||| ||  |   |||     ||||  |||
Db    528 GGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCC 469

Qy   1935 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 1994
          | || || || | |||||| ||  || |  ||   ||| || |     | || |||||
Db    468 ACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCG 409

Qy   1995 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 2054
          |   || || || || || || ||||| || || |||||||| || ||| | || ||||
Db    408 CCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAA 349

Qy   2055 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 2114
          | || |||||  |  ||   |  ||  | || || || |||||||| ||| || |||| |
Db    348 AACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGA 289

Qy   2115 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 2174
          |||| |||||    || |  | |||| | || || || |||||||| | |  |||  |||
Db    288 ACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCA 229

Qy   2175 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 2234
            ||||| ||  | ||| |   |     | ||||||||  |  |  | |||||||  || |
Db    228 AGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATT 169

Qy   2235 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 2294
          || |   |  ||||    |       ||        |  ||||       |   | |||
Db    168 TTCANGTGGCGAAGACACTTTAAAGCAAGGGCAGTGGCATTTGCCCGAGGC-CACCTGTC 110

Qy   2295 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAA 2353
          |||||| ||| | || |  |||| | |  |||| | ||| |   |  |  ||| | ||
Db    109 AGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAG 50

Qy   2354 TGGTTTTGGAATTAA 2368
          ||||  |||  || |
Db     49 TGGTACTGGGGTTGA 35
```

RESULT 10
BB595164

```
LOCUS       BB595164                 648 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB595164 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION   BB595164
VERSION     BB595164.2  GI:16450167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11491766.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C030004N01"
                     /clone_lib="RIKEN full-length enriched, adult male corpus
                     striatum"
                     /sex="male"
                     /tissue_type="corpus striatum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      141 a    206 c    138 g    163 t
ORIGIN
```

```
  Query Match             10.4%;  Score 292.6;  DB 9;  Length 648;
  Best Local Similarity   90.2%;  Pred. No. 2.8e-33;
  Matches  313;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          ||||||||||||||||||||||||||||| ||||||||||||||| |||| |||||||| |
Db    302 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 361

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| |||||||||||||| |||||||| ||||| || ||||||||||||| |||
Db    362 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 421

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||||| || || || |||||||||||||||||||| ||||| ||||| ||||
```

```
Db     422 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 481

Qy    1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
           |||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||
Db     482 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 541

Qy    1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
           |||||||||||||||||||||||||||| |||| ||| |||||||||||||| ||||| |
Db     542 ACCTTACATCCTGCCAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAACAGACCGCT 601

Qy    1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATC 1627
           || |||||||||||||||||| || ||||| || ||   |||||||
Db     602 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATATAGACATC 648


RESULT 11
BB649967
LOCUS       BB649967                625 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB649967 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION   BB649967
VERSION     BB649967.1  GI:16484222
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C130099N17"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="16 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT      136 a    198 c    135 g    156 t
ORIGIN

  Query Match             9.9%;  Score 280.4;  DB 9;  Length 625;
  Best Local Similarity   91.9%;  Pred. No. 1.7e-31;
  Matches  296;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy    1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
           |||||||||||||||||||||||||||| |||||||||||||| |||| |||||||| |
Db     303 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 362

Qy    1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
           |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db     363 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 422

Qy    1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
           |||||||||||| || || || |||||||||||||||||| ||||| ||||| ||||
Db     423 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 482

Qy    1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
           |||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||
Db     483 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 542

Qy    1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
           |||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |
Db     543 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 602

Qy    1581 CACTGTACAACAAATCCCATGC 1602
           || ||||||||||||||||||
Db     603 CATTGTACAACAAATCCCATGC 624


RESULT 12
BB640911
LOCUS       BB640911                767 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB640911 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730093G08 5', mRNA sequence.
ACCESSION   BB640911
VERSION     BB640911.1  GI:16476069
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
```

,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .767
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730093G08"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 175 a 239 c 176 g 172 t 5 others

ORIGIN

Query Match 9.6%; Score 269.6; DB 9; Length 767;
Best Local Similarity 88.3%; Pred. No. 5.7e-30;

Matches 293; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          |||||||||||||||||||||||||||||| ||||||||||||||| |||| |||||||| |
Db    339 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 398

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| |||||||||||||| |||||||| ||||| || ||||||||||| |||
Db    399 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 458

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||||| || || || |||| |||||||||||||| ||||| ||||| ||||
Db    459 TCAGTCCAATGGAGTCTGCAAGTTCTGGGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 518

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          |||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||
Db    519 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 578

Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580
          |||||||||||| |||||||||||||||| ||||| ||  ||||||| |||||  | || |
Db    579 ACCTTACATCCGGCCAAGCAATGTTCAATATCCAGGGGGGCCATCTTACAACAAAACGGT 638

Qy   1581 CACTGTACAACAAATCCCATGCAGTTCTGATC 1612
          || || |||||||||||||  |  |  ||||||
Db    639 CATTGGACAACAAATCCCAGGGGGGCCTGATC 670
```

RESULT 13
BM021317
LOCUS BM021317 568 bp mRNA linear EST 30-OCT-2001
DEFINITION ie75f03.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA 5' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;, mRNA sequence.
ACCESSION BM021317
VERSION BM021317.1 GI:16535673
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 568)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T. , Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Other_ESTs: ie75f03.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
High quality sequence stop: 436.
FEATURES Location/Qualifiers
source 1. .568
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;

Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 127 a 170 c 123 g 148 t
ORIGIN

Query Match 8.9%; Score 251; DB 10; Length 568;
Best Local Similarity 100.0%; Pred. No. 3.1e-27;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    318 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 377

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 437

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 497

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 557

Qy   1521 ACCTTACATCC 1531
          ||||||||||||
Db    558 ACCTTACATCC 568
```

RESULT 14
BB651409
LOCUS BB651409 628 bp mRNA linear EST 26-OCT-2001
DEFINITION BB651409 RIKEN full-length enriched, 0 day neonate cerebellum Mus musculus cDNA clone C230079N02 5', mRNA sequence.
ACCESSION BB651409
VERSION BB651409.1 GI:16485662
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 628)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .628
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230079N02"
/clone_lib="RIKEN full-length enriched, 0 day neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 129 a 200 c 143 g 152 t 4 others
ORIGIN

Query Match 8.1%; Score 227.6; DB 9; Length 628;
Best Local Similarity 88.4%; Pred. No. 7.4e-24;
Matches 245; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy   1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340
          |||||||||||||||||||||||||| |||||||||||||| |||| |||||||| |
Db    339 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 398

Qy   1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400
          |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db    399 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 458

Qy   1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460
          |||||||||||||  ||  ||  |||||||||||||||||||| ||||| ||||| ||||
Db    459 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 518

Qy   1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520
          ||||||||  |||| ||||||||  |||||||  | |||  |||||||  |||||||||||  ||
Db    519 CTGATTATGTTAACGAAACTACCATTCTGGNTTAGGTATGGCCATNTGGGCAGACCTNTG 578
```

```
Qy   1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGG 1557
          ||||||||||||||||||   |||||  | ||||||||
Db    579 ACCTTACATCCTGCCAAAGCATGTTTCATATCCAAGG 615


RESULT 15
BM082484/c
LOCUS       BM082484                 570 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223
            3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;,
            mRNA sequence.
ACCESSION   BM082484
VERSION     BM082484.1  GI:16929414
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN

  Query Match             6.5%;  Score 184.2;  DB 10;  Length 570;
  Best Local Similarity   59.2%;  Pred. No. 1.5e-17;
  Matches  337;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    1;

Qy   1670 TTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGC 1729
           |||||||| |||||||||  |  |||| ||||| || ||  |||||     ||    |||
Db    570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy   1730 ATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCT 1789
           | ||  || |||| ||||  |||||  |  ||| | |  ||| | ||||| | ||| |
Db    510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy   1790 TATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAA 1849
                | ||    ||    |  | |    ||   |  ||||   | || |  || || |||
Db    450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy   1850 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGAGTCAAG 1900
           | || || ||| ||||||||||||     |||          ||     || ||   || |
Db    390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331

Qy   1901 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 1960
            |||    ||  |||||  |  |||  |  | ||    | |||| | || | |  |  |
Db    330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271

Qy   1961 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 2020
           ||| |  |  |||||  | |   |||||||  |  ||   || |||||||| || |||||
Db    270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211

Qy   2021 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 2080
           || || || |||||||||| |||||||| || ||| |  |   |  || ||  || ||
Db    210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151

Qy   2081 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 2140
           |||| ||||| ||| |||| ||  | ||| ||||||||||| || |  |||||  | |||
Db    150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91

Qy   2141 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 2200
           | || |||||||||  | || |       |  |||     ||||||  ||| || |
Db     90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31

Qy   2201 GTCGACAAAAACAATAAGTTATACCTTAG 2229
           |  | |||||  | |     ||| |  ||
Db     30 CTAAATAAAAATGACATTGCATATCACAG 2


Search completed: October  6, 2002, 07:15:52
Job time : 2383.78 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 57.5794 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-4
Perfect score: 2822
Sequence: 1 aagctttgctactcagagag.....aacatgttgcgattgaattc 2822

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	4.5	1814	2	US-08-483-151-1	Sequence 1, Appli
2	127	4.5	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	113.4	4.0	1654	1	US-07-914-281-13	Sequence 13, Appl
4	113.4	4.0	1654	1	US-08-393-246-13	Sequence 13, Appl
5	113.4	4.0	1654	1	US-08-525-058A-13	Sequence 13, Appl
6	113.4	4.0	1654	2	US-08-696-731-13	Sequence 13, Appl
7	113.4	4.0	1654	4	US-09-042-531-13	Sequence 13, Appl
8	112.2	4.0	1086	1	US-07-914-281-12	Sequence 12, Appl
9	112.2	4.0	1086	1	US-08-393-246-12	Sequence 12, Appl
10	112.2	4.0	1086	1	US-08-525-058A-12	Sequence 12, Appl
11	112.2	4.0	1086	2	US-08-696-731-12	Sequence 12, Appl
12	112.2	4.0	1086	4	US-09-042-531-12	Sequence 12, Appl
13	112.2	4.0	2042	3	US-09-063-237-2	Sequence 2, Appli
14	112.2	4.0	2043	1	US-07-914-281-1	Sequence 1, Appli
15	112.2	4.0	2043	1	US-08-393-246-1	Sequence 1, Appli
16	112.2	4.0	2043	1	US-08-273-411-4	Sequence 4, Appli
17	112.2	4.0	2043	1	US-08-525-058A-1	Sequence 1, Appli
18	112.2	4.0	2043	2	US-08-696-731-1	Sequence 1, Appli
19	112.2	4.0	2043	4	US-09-042-531-1	Sequence 1, Appli
20	112.2	4.0	2043	5	PCT-US91-00899-1	Sequence 1, Appli
21	111.8	4.0	1316	1	US-07-914-281-10	Sequence 10, Appl
22	111.8	4.0	1316	1	US-08-393-246-10	Sequence 10, Appl
23	111.8	4.0	1316	1	US-08-525-058A-10	Sequence 10, Appl
24	111.8	4.0	1316	2	US-08-696-731-10	Sequence 10, Appl
25	111.8	4.0	1316	4	US-09-042-531-10	Sequence 10, Appl
26	100.8	3.6	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	100.8	3.6	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	100.8	3.6	1488	1	US-07-914-281-9	Sequence 9, Appli
29	100.8	3.6	1488	1	US-08-393-246-9	Sequence 9, Appli
30	100.8	3.6	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	100.8	3.6	1488	2	US-08-696-731-9	Sequence 9, Appli
32	100.8	3.6	1488	4	US-09-042-531-9	Sequence 9, Appli
33	100.8	3.6	2175	4	US-08-482-073-9	Sequence 9, Appli
34	100.8	3.6	2861	4	US-08-482-073-10	Sequence 10, Appl
35	100.8	3.6	3647	1	US-07-914-281-7	Sequence 7, Appli
36	100.8	3.6	3647	1	US-08-393-246-7	Sequence 7, Appli
37	100.8	3.6	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	100.8	3.6	3647	2	US-08-696-731-7	Sequence 7, Appli
39	100.8	3.6	3647	4	US-09-042-531-7	Sequence 7, Appli
40	100.8	3.6	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	97.6	3.5	2134	2	US-08-483-151-3	Sequence 3, Appli
42	97.6	3.5	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	63	2.2	4673	1	US-07-638-431-1	Sequence 1, Appli
c 44	63	2.2	4673	5	PCT-US92-00018-1	Sequence 1, Appli
45	54.2	1.9	2042	2	US-08-933-821-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 4.5%; Score 127; DB 2; Length 1814;
Best Local Similarity 52.3%; Pred. No. 1.4e-18;
Matches 330; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

```
Qy   1557 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 1616
          | |||| ||| |   |  ||||   |||     ||   |  |||| ||  |  |||| |
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy   1617 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 1676
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy   1677 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 1736
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy   1737 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 1796
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy   1797 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 1856
Db    788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy   1857 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 1916
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy   1917 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 1976
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy   1977 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 2036
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy   2037 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 2093
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy   2094 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 2153
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy   2154 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;    APPLICANT:  The General Hospital Corporation
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             4.5%;  Score 127;  DB 5;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 1.4e-18;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy   1557 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 1616
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy   1617 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 1676
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy   1677 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 1736
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy   1737 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 1796
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy   1797 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 1856
Db    788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy   1857 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 1916
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy   1917 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 1976
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy   1977 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 2036
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy   2037 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 2093
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy   2094 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 2153
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy   2154 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 2184
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 3
US-07-914-281-13
; Sequence 13, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-13

  Query Match             4.0%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.1e-15;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
          || ||| ||||| |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
           |      ||      |  ||| |  |||| | |||||     ||||      |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
          ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
          |||     ||| || ||||          | | |  |     |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
           |  |  |   |||      |  |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
          |||||  ||    || ||| |    ||           |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
           |     |     |     |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
          |   | ||| | ||  | |||| || |||| || ||  |   |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
            ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   2246 AAG 2248
          ||
Db   1095 GAG 1097


RESULT 4
US-08-393-246-13
; Sequence 13, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-13

  Query Match             4.0%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.1e-15;
```

```
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
          || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
           |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
          || ||  |  | || |||||||  |||  |  | |  |   |  | | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
          ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
          |||     ||| || ||||          | | | |       |    | |      | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
           |  |  |   |||      | |||||  |||||  |||   ||||||  ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
          |||||  ||    || ||| |    ||          |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGCTACGG---A 794

Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
             |    |    |    |  ||  ||||   ||  |  ||    |  ||  ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
          |||||  |||| || || ||   | |||   || ||||||||  || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
          ||  |   |||  |  | || || || || ||||| ||     ||     ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
          |   | ||| | ||  | ||||| || ||||| || |    |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
            ||| ||| |||| | |||||  || |       ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   2246 AAG 2248
           ||
Db    1095 GAG 1097


RESULT 5
US-08-525-058A-13
; Sequence 13, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-13

  Query Match             4.0%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.1e-15;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
          || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
           |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
          || ||  |  | || |||||||  |||  |  | |  |   |  | | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
          ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
          |||     ||| || ||||          | | | |       |    | |      | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
           |  |  |   |||      | |||||  |||||  |||   ||||||  ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
          |||||  ||    || ||| |    ||          |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGCTACGG---A 794

Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
             |    |    |    |  ||  ||||   ||  |  ||    |  ||  ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
          |||||  |||| || || ||   | |||   || ||||||||  || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914
```

```
Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
          |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
            ||| ||| ||||  | ||||  || |       ||||| || |||||| ||||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   2246 AAG 2248
            ||
Db   1095 GAG 1097
```

```
RESULT 6
US-08-696-731-13
; Sequence 13, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-13

  Query Match             4.0%;  Score 113.4;  DB 2;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.1e-15;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;
```

```
Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
          || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
           |      ||     | ||| |  |||| | |||||      ||||     |  | |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
          || ||  |  | ||  |||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
          ||   ||  | ||| ||    |  || || || ||  || | ||||||  | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
          |||     ||| || ||||           | | | |       |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
           |  |  |   |||      | |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
          |||||  ||    || ||| |   ||          |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
             |      |     |     |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
          |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
            ||| ||| ||||  | ||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   2246 AAG 2248
            ||
Db   1095 GAG 1097
```

```
RESULT 7
US-09-042-531-13
; Sequence 13, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
```

```
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/042,531
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;     INFORMATION FOR SEQ ID NO:  13:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH:  1654 base pairs
;         TYPE:  nucleic acid
;         STRANDEDNESS:  unknown
;         TOPOLOGY:  unknown
;       MOLECULE TYPE:  DNA (genomic)
US-09-042-531-13

 Query Match
 Best Local Similarity   51.2%;  Score 113.4;  DB 4;  Length 1654;
 Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy   1484 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 1543
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy   1544 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 1600
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy   1601 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 1657
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy   1658 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 1717
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy   1718 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 1777
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy   1778 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy   1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy   1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

 Qy   1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008
 Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

 Qy   2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 2065
 Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

 Qy   2066 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 2125
 Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

 Qy   2126 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 2185
 Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

 Qy   2186 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 2245
 Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

 Qy   2246 AAG 2248
 Db   1095 GAG 1097

RESULT 8
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;    INFORMATION FOR SEQ ID NO:  12:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  1086 base pairs
;        TYPE:  NUCLEIC ACID
;        STRANDEDNESS:  unknown
;        TOPOLOGY:  unknown
;      MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

 Query Match
 Best Local Similarity   51.6%;  Score 112.2;  DB 1;  Length 1086;
 Matches  364;  Conservative    0;  Mismatches  323,  Indels   18,  Gaps    4,
```

```
Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
            | || ||||  |   |   |  | ||    | || |       | |||||  |      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
             ||||| ||| | || ||||| ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |   ||  |   | |   |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
           |    ||            |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          ||  |  |  |   ||||    ||  | ||    |  || || || ||  ||||  || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || ||||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    |||||  |||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |      |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 9
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             4.0%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 1.8e-15;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
            | || ||||  |   |   |  | ||    | || |       | |||||  |      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
             ||||| ||| | || ||||| ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |   ||  |   | |   |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
           |    ||            |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          ||  |  |  |   ||||    ||  | ||    |  || || || ||  ||||  || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || ||||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    |||||  |||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
```

```
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 10
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
;   APPLICANT:  LOWE, JOHN B.
;   TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;   TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;   TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;   TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;     STREET:  1755 Jefferson Davis Highway, Fourth Floor
;     CITY:  Arlington
;     STATE:  Virginia
;     COUNTRY:  U.S.A.
;     ZIP:  22202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/525,058A
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Lavalleye, Jean-Paul M. P.
;     REGISTRATION NUMBER:  31,451
;     REFERENCE/DOCKET NUMBER:  2363-060-55
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)521-4500
;     TELEFAX:  (703)486-2347
;     TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1086 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

 Query Match             4.0%;  Score 112.2;  DB 1;  Length 1086;
 Best Local Similarity   51.6%;  Pred. No. 1.8e-15;
 Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618

Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675

Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735

Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795

Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT--------GTTTGAAGTGCCAAGCAAAGAG 1846

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906

Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966

Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026

Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083

Db    748 TCCTTGCACCCCGACTCATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 11
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
;   APPLICANT:  LOWE, JOHN B.
;   TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;   TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;   TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;   TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;   NUMBER OF SEQUENCES:  14
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;     ADDRESSEE:  P.C.
;     STREET:  1755 Jefferson Davis Highway, Fourth Floor
;     CITY:  Arlington
;     STATE:  Virginia
;     COUNTRY:  U.S.A.
;     ZIP:  22202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/696,731
;     FILING DATE:  14-AUG-1996
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/393,246
;     FILING DATE:
;     APPLICATION NUMBER:  US 08/220,433
;     FILING DATE:  30-MAR-1994
;     APPLICATION NUMBER:  US 07/914,281
;     FILING DATE:  20-JUL-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Lavalleye, Jean-Paul M. P.
;     REGISTRATION NUMBER:  31,451
;     REFERENCE/DOCKET NUMBER:  2363-060-55
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)521-4500
;     TELEFAX     (703)486-2347
```

```
;      TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1086 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             4.0%;  Score 112.2;  DB 2;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 1.8e-15;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          ||||| |||| | ||||| |||||  |  | | | || | | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
          | || |||| |  |   | | ||   | || |     | ||||| |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
          ||||| ||| | || |||||  |||| || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |  ||  |   |  |   |       |         | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
          |  |||||  |||||  |||   |||||||| || ||     |||||  ||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
          |   ||           |  |  |  |    |||||  |   |                 |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || |||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || |||||.||    ||    ||||| |||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| |||| |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 12
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1086 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match             4.0%;  Score 112.2;  DB 4;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 1.8e-15;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          ||||| |||| | ||||| |||||  |  | | | || | | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
          | || |||| |  |   | | ||   | || |     | ||||| |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
          ||||| ||| | || |||||  |||| || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |  ||  |   |  |   |       |         | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
          |  |||||  |||||  |||   |||||||| || ||     |||||  ||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
          |   ||           |  |  |  |    |||||  |   |                 |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690
```

```
Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |      ||||| || |||||| |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 13
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             4.0%;  Score 112.2;  DB 3;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 2e-15;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
          | || ||||  |   |   |  | ||    | || |       | |||||  |      |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
          |||||  ||| | || |||||  ||||  || ||   |  ||    |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
          |  || || || ||  || | ||||||  | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |   ||  |   |  |     |      |         | |  |  |  |  |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
          |    ||           |  |  |  |    |||||  |   |                |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |      ||||| || |||||| |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             4.0%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 2e-15;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
            | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
           ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
          |   ||  |   | |    |     |          | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966
           |    ||           |  |  |  |    |||||  |   |               |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083
          ||  | |||   || ||||||| || ||||| |     || ||  |  |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143
          || || || || ||||| ||    ||    |||||  |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203
          |||||  || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248
          |||||  || |       |||||  || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 15
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             4.0%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 2e-15;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy   1559 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 1618
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy   1619 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 1675
            | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy   1676 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 1735
           ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy   1736 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 1795
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy   1796 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 1846
```

```
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy   1847 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 1906

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy   1907 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 1966

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy   1967 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 2026

Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy   2027 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 2083

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy   2084 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 2143

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy   2144 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 2203

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   2204 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:21:20
Job time : 103.579 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 1353.36 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-4_COPY_1289_2368
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1080	100.0	3019	9	AB023021	AB023021 Homo sapi
2	1078.4	99.9	80247	9	AL512406	AL512406 Human DNA
3	1075.2	99.6	2501	9	HSA238701	AJ238701 Homo sapi
4	931.2	86.2	2139	10	AB015426	AB015426 Mus muscu
5	929.6	86.1	1128	10	AF345993	AF345993 Rattus no
6	928	85.9	2156	10	AB049819	AB049819 Rattus no
7	916.8	84.9	1705	10	AF230460	AF230460 Cricetulu
8	783.6	72.6	1707	5	AB035906	AB035906 Gallus ga
9	676.8	62.7	1080	5	AB035905	AB035905 Xenopus l
10	639	59.2	71396	2	AC100674	AC100674 Mus muscu
11	277.8	25.7	1567	5	AB023627	AB023627 Danio rer
12	234.6	21.7	1832	5	AB023628	AB023628 Danio rer
13	222.6	20.6	1092	5	AB035907	AB035907 Xenopus l
14	220.2	20.4	1264	5	AF288369	AF288369 Gallus ga
15	176.2	16.3	2657	5	GGU73678	U73678 Gallus gall
16	145.2	13.4	1429	10	CGU78737	U78737 Cricetulus
17	145.2	13.4	1462	10	AF090450	AF090450 Cricetulu
18	143.6	13.3	1718	10	AF090449	AF090449 Cricetulu
19	137.2	12.7	1044	5	AB035908	AB035908 Gallus ga
20	132	12.2	2293	4	BTA134FTG	X87810 B.taurus al
21	132	12.2	12607	4	BTA132772	AJ132772 Bos tauru
c 22	128	11.9	71396	2	AC100674	AC100674 Mus muscu
23	127.6	11.8	1158	10	AB039126	AB039126 Mus muscu
24	127	11.8	1487	3	AF016899	AF016899 Schistoso
25	127	11.8	1814	6	AR028686	AR028686 Sequence
26	127	11.8	3594	10	MMU45980	U45980 Mus musculu
c 27	127	11.8	117993	2	AC091284	AC091284 Mus muscu
28	125.6	11.6	1613	4	BTA132776	AJ132776 Bos tauru
29	125.6	11.6	1619	4	BTA132775	AJ132775 Bos tauru
30	125.6	11.6	1644	4	BTA132774	AJ132774 Bos tauru
31	125.6	11.6	1873	4	BTA132773	AJ132773 Bos tauru
32	124.6	11.5	1119	9	AF345881	AF345881 Macaca mu
33	122.8	11.4	1170	10	AB039127	AB039127 Mus muscu
34	122.8	11.4	1170	10	AB039131	AB039131 Mus muscu
35	122.8	11.4	1615	10	MMU33457	U33457 Mus musculu
36	122	11.3	1113	9	AF345884	AF345884 Saimiri s
37	121.2	11.2	1170	10	AB039124	AB039124 Mus muscu
38	121.2	11.2	1170	10	AB039125	AB039125 Mus muscu
39	121.2	11.2	1170	10	AB039128	AB039128 Mus muscu
40	121.2	11.2	1170	10	AB039129	AB039129 Mus muscu
41	121.2	11.2	1170	10	AB039130	AB039130 Mus muscu
42	121.2	11.2	1170	10	AB039132	AB039132 Mus muscu
43	119.8	11.1	1119	9	AF345883	AF345883 Macaca mu
44	118.4	11.0	1170	10	AB039133	AB039133 Mus spici
45	118.2	10.9	1080	9	AB051859	AB051859 Pongo pyg

ALIGNMENTS

RESULT 1
AB023021
LOCUS AB023021 3019 bp mRNA linear PRI 23-JUN-1999
DEFINITION Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX, complete cds.
ACCESSION AB023021
VERSION AB023021.1 GI:5139692
KEYWORDS alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S., Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and Narimatsu,H.
TITLE Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX
JOURNAL FEBS Lett. 453, 237-242 (1999)
REFERENCE 2 (bases 1 to 3019)
AUTHORS Kaneko,M., Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
ORIGIN

  Query Match             100.0%;  Score 1080;  DB 9;  Length 3019;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-268;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 654

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 714

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 774

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 1074

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1134

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1194

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1195 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1255 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1314

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1374


RESULT 2
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
```

of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-504J9 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP3-381A2 is at 80148 in this sequence.
The true right end of clone RP11-77B15 is at 100 in this sequence.

FEATURES	Location/Qualifiers
source	1. .80247
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="RP11-504J9"
	/clone_lib="RPCI-11.2"
repeat_region	184. .1422
	/note="L2 repeat: matches 875. .2419 of consensus"
repeat_region	1744. .1834
	/note="MIR repeat: matches 166. .258 of consensus"
repeat_region	2950. .2977
	/note="14 copies 2 mer ca 100% conserved"
repeat_region	3411. .3669
	/note="LTR16C repeat: matches 93. .364 of consensus"
repeat_region	4372. .4704
	/note="MLT1A1 repeat: matches 5. .365 of consensus"
misc_feature	complement(7821. .8406)
	/note="match: GSS: Em:AQ550748"
repeat_region	8126. .8184
	/note="MER58B repeat: matches 272. .332 of consensus"
repeat_region	8185. .8696
	/note="MLT2D repeat: matches 1. .513 of consensus"
repeat_region	8779. .8817
	/note="MLT2D repeat: matches 504. .542 of consensus"
repeat_region	8819. .8974
	/note="MER58A repeat: matches 6. .162 of consensus"
misc_feature	complement(9443. .9891)
	/note="match: GSS: Em:B90633"
misc_feature	complement(9452. .9883)
	/note="match: GSS: Em:AQ119695"
repeat_region	9782. .9894
	/note="MIR repeat: matches 92. .205 of consensus"
repeat_region	10089. .10212
	/note="MIR repeat: matches 135. .262 of consensus"
misc_feature	complement(10093. .10671)
	/note="match: GSS: Em:AQ237061"
repeat_region	10223. .10282
	/note="MIR repeat: matches 203. .262 of consensus"
misc_feature	complement(10315. .10631)
	/note="match: GSS: Em:AQ200259"
misc_feature	10805. .11186
	/note="match: GSS: Em:AQ072353"
repeat_region	11657. .11844
	/note="L2 repeat: matches 2546. .2750 of consensus"
misc_feature	complement(11659. .12084)
	/note="match: GSS: Em:AQ518361"
misc_feature	complement(11674. .11987)
	/note="match: GSS: Em:AQ094678"
misc_feature	12083. .12588
	/note="match: GSS: Em:AQ691235"
repeat_region	12596. .12810
	/note="MIR repeat: matches 23. .248 of consensus"
repeat_region	12920. .12986
	/note="L1MA5 repeat: matches 6228. .6295 of consensus"
repeat_region	12987. .13169
	/note="L1MA5 repeat: matches 5852. .6043 of consensus"
repeat_region	13168. .13280
	/note="L1MA8 repeat: matches 5710. .5832 of consensus"
repeat_region	13851. .17218
	/note="L1PB1 repeat: matches 2787. .6155 of consensus"
repeat_region	18903. .18953
	/note="MLT2CB repeat: matches 439. .501 of consensus"
repeat_region	18984. .19433
	/note="MLT2CB repeat: matches 1. .460 of consensus"
misc_feature	complement(19108. .19690)
	/note="match: GSS: Em:AQ563935"
repeat_region	20683. .20801
	/note="AluJo/FRAM repeat: matches 169. .288 of consensus"
repeat_region	20802. .20862
	/note="AluJ/FLAM repeat: matches 3. .66 of consensus"
repeat_region	20912. .21033
	/note="MIR repeat: matches 114. .237 of consensus"
repeat_region	21047. .21128
	/note="41 copies 2 mer ta 63% conserved"
repeat_region	21676. .21859
	/note="L2 repeat: matches 2130. .2308 of consensus"
repeat_region	21890. .23612
	/note="L1P repeat: matches 841. .2570 of consensus"
repeat_region	23601. .26609
	/note="L1PA5 repeat: matches 3107. .6145 of consensus"
repeat_region	26620. .26789
	/note="MER5A repeat: matches 1. .174 of consensus"
repeat_region	28521. .28627
	/note="MER39 repeat: matches 5. .111 of consensus"
repeat_region	28656. .28837
	/note="MER39b repeat: matches 54. .235 of consensus"
repeat_region	28838. .29142
	/note="AluSc repeat: matches 4. .308 of consensus"
repeat_region	29143. .29494
	/note="MER39b repeat: matches 235. .549 of consensus"
repeat_region	30529. .31344
	/note="L1PA3 repeat: matches 5331. .6146 of consensus"
repeat_region	32234. .32340
	/note="L2 repeat: matches 2627. .2750 of consensus"
repeat_region	32399. .32578
	/note="MIR repeat: matches 12. .203 of consensus"
repeat_region	32816. .32979
	/note="MIR repeat: matches 91. .259 of consensus"
misc_feature	complement(33006. .33450)
	/note="match: GSS: Em:AQ261227"
misc_feature	complement(33058. .33444)
	/note="match: GSS: Em:AQ040453"
repeat_region	33312. .33596
	/note="L1M4 repeat: matches 3473. .3773 of consensus"
repeat_region	34856. .34915
	/note="30 copies 2 mer aa 68% conserved"
misc_feature	complement(34987. .35547)
	/note="match: GSS: Em:AQ321598
	match: STS: Em:G55571"
misc_feature	complement(36488. .36936)
	/note="match: GSS: Em:B41073"
misc_feature	complement(36654. .37090)
	/note="match: GSS: Em:AQ772584"
misc_feature	37102. .37591
	/note="match: GSS: Em:AQ489567"
misc_feature	complement(37581. .38221)
	/note="match: GSS: Em:AQ418233"
repeat_region	37720. .37772
	/note="L2 repeat: matches 2657. .2710 of consensus"
misc_feature	complement(37778. .38215)
	/note="match: GSS: Em:AQ819773"
misc_feature	38218. .38732
	/note="match: GSS: Em:AQ599868"
repeat_region	38861. .39164
	/note="AluSx repeat: matches 1. .305 of consensus"
repeat_region	39612. .39653
	/note="21 copies 2 mer ac 90% conserved"
repeat_region	40854. .40909
	/note="L2 repeat: matches 2655. .2710 of consensus"
repeat_region	42169. .42270
	/note="MER5A repeat: matches 1. .111 of consensus"
repeat_region	42411. .42548
	/note="MER5B repeat: matches 21. .181 of consensus"
repeat_region	42780. .42938
	/note="MER5B repeat: matches 17. .177 of consensus"
repeat_region	43229. .43326
	/note="MIR repeat: matches 40. .148 of consensus"

```
     repeat_region   45383. .45428
                     /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                     /note="match: STS: Em:G51995
                     match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                     /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
                     /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                     /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                     /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match             99.9%;  Score 1078.4;  DB 9;  Length 80247;
  Best Local Similarity   99.9%;  Pred. No. 3e-268;
  Matches 1079;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35559 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 35618

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35619 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 35678

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35679 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 35738

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35739 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 35798

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35799 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 35858

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35859 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 35918

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35919 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 35978

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35979 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 36038

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36039 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 36098

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36099 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 36158

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36159 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 36218

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   36219 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTACTTGTAAATTT 36278

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36279 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 36338

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   36339 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 36398

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36399 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 36458

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36459 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 36518

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36519 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 36578

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36579 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 36638


RESULT 3
HSA238701
LOCUS       HSA238701              2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
```

```
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match             99.6%;  Score 1075.2;  DB 9;  Length 2501;
  Best Local Similarity   99.7%;  Pred. No. 2.2e-267;
  Matches 1077;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 164

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 224

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 284

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 344

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 404

Qy    301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          |||||||||||||||||||||||||||||||||||| |||||||||| || |||
Db    405 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGCTGGGATCTGACTAACTTA 464

Qy    361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 524

Qy    421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 584

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    585 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 644

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    645 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 704

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    705 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 764

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    765 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 824

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    825 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 884

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    885 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 944

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    945 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1004

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1005 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1064

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1065 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1124

Qy   1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1125 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1184


RESULT 4
AB015426
LOCUS       AB015426                2139 bp    mRNA    linear   ROD 23-JUN-1999
DEFINITION  Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete
            cds.
ACCESSION   AB015426
VERSION     AB015426.1  GI:3702718
KEYWORDS    Fut9; alpha1,3-fucosyltransferase IX.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K.
            and Narimatsu,H.
  TITLE     Expression cloning and characterization of a novel murine alpha1,
            3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15)
            epitope in brain and kidney
  JOURNAL   J. Biol. Chem. 273 (41), 26729-26738 (1998)
  MEDLINE   98434588
REFERENCE   2  (bases 1 to 2139)
  AUTHORS   Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN

  Query Match             86.2%;  Score 931.2;  DB 10;  Length 2139;
  Best Local Similarity   91.4%;  Pred. No. 3.7e-230;
  Matches  987;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          |||||||||||||||||| |||||||||||||| |||| |||||||| ||||||||||
Db    115 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 174

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    175 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 234
```

```
Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||| || || || ||||||||||||||||||||| ||||| ||||| ||||||||||||
Db    235 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 294

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| ||||||| ||||||| ||||| ||||||||||||||||||||||||||||||||
Db    295 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 354

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db    355 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 414

Qy    301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          |||||||||||||| || ||||| || ||  |||||||||| |||||||||| || |||
Db    415 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 474

Qy    361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          |||||||| || |||||||||| |||||||||||||||||||||| || ||||| |||
Db    475 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 534

Qy    421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||| |||||||||||||||||| |||||||| |||||||||| ||||| ||||||
Db    535 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 594

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db    595 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 654

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db    655 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 714

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db    715 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 774

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          ||||| |||||||| || || |||||||| ||||||| |||||||| ||||||||||
Db    775 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 834

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          |||||||| |||||||| ||||| ||||| |||||||||| ||||||| ||||||||
Db    835 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 894

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||||||||||
Db    895 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 954

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          |||||||| |||||||||||||||||||||||| |||||||||||||| |||||||
Db    955 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 1014

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          |||||||||||||| |||||||||||||||| |||||||||||||||||||||||| |||
Db   1015 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||| |||||||| || ||||| ||||||||||||||||||||  |||| ||||| |||
Db   1075 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1134

Qy   1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1135 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194
```

```
RESULT 5
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN
```

```
  Query Match           86.1%;  Score 929.6;  DB 10;  Length 1128;
  Best Local Similarity   91.3%;  Pred. No. 9.8e-230;
  Matches  986;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||| |||||||||||||| |||||||||||||| |||| ||||||||||||||||||
Db     49 ATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGGGC 108

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          || ||| |||||||||| || || |||||||| || |||||||||| ||||||||||
Db    109 TGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 168

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||| || || |||||||||||||||||||||| ||||| ||||  ||||||||||||
Db    169 ATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATTAT 228

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 TTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 288

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          |||||||||||| |||||||||||||||| |||||||||||||||||||||  |||||
Db    289 TCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGTAC 348

Qy    301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          |||||||||||||| || ||||| || ||| |||||||||| |||||||||| || |||
Db    349 AACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTA 408

Qy    361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          |||||||| || ||||| ||||| |||||||||||||||||||| || || ||||| |||
Db    409 CCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCACT 468
```

```
Qy    421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||| ||||||||||||||||||||||||||||| |||||||||| ||||| ||||||
Db    469 CACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCCGT 528

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||||||||| |||||||||| |||||||||||||| ||||||  ||||| ||||||
Db    529 GATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTT 588

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||||||| || |||||||||||||||||||||||| || ||||||
Db    589 GAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCAT 648

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| ||||||
Db    649 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGCAA 708

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          |||||||||||||||| |||||||||||| ||||||| ||||||||| ||||||||||
Db    709 GCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 768

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          |||||||||||||||| ||||| ||||| |||||||||||| |||||||| ||||||||
Db    769 TATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 828

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||| ||||||| |||||||||||||| ||||| ||||||||||||||||||||||||||
Db    829 TTTTTGGCTGGTTCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 888

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          |||||||| |||||||||||||||||||||||| |||||||||||||| | |||||
Db    889 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTGGCAAAA 948

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          |||||||||| || |||||||||||||||||| |||||||||| ||||||||||||||||
Db    949 TATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGAAG 1008

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||| |||||||| || ||||| |||||||||||||||||||  |||| |||||||||
Db   1009 GATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATCAT 1068

Qy   1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          || |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1069 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1128
```

RESULT 6
AB049819

```
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN
```

```
  Query Match             85.9%;  Score 928;  DB 10;  Length 2156;
  Best Local Similarity   91.2%;  Pred. No. 2.5e-229;
  Matches  985;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||| |||||||||||||| |||||||||||||| |||| ||||||||||||||||||
Db    346 ATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGGGC 405

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          || ||| |||||||||| || || |||||||| || |||||||||||| ||||||||||
Db    406 TGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 465

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||| || || |||||||||||||||||||||| ||||| ||||  ||||||||||||
Db    466 ATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATTAT 525

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    526 TTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 585

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          |||||||||| |||||||||||||||| |||||||||||||||||||| ||  |||||
Db    586 TCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGTAC 645

Qy    301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          ||||||||||||| || ||||| || ||| |||||||||| |||||||||||| || |||
Db    646 AACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTA 705

Qy    361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          |||||||| || ||||| ||||| |||||||||||||||||||| || || ||||| |||
Db    706 CCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCACT 765

Qy    421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||| ||||||||||||||||||||||||||||| |||||||||| ||||| ||||||
Db    766 CACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCCGT 825

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||||||||| |||||||||| |||||||||||||| |||||||  ||||| ||||||
Db    826 GATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTT 885

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |||||||||||||||||| || |||||||||||||||||||||||| || ||||||
Db    886 GAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCAT 945

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| ||||||
Db    946 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGCAA 1005

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          |||||||||||||||| |||||||||||| ||||||| ||||||||| ||||||||||
```

```
Db   1006 GCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 1065

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780

Db   1066 TATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 1125

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840

Db   1126 TTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1185

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900

Db   1186 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGTGAGTTGGCAAAA 1245

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960

Db   1246 TATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGAAG 1305

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020

Db   1306 GATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATCAT 1365

Qy   1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080

Db   1366 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1425


RESULT 7
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                                CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN

  Query Match             84.9%;  Score 916.8;  DB 10;  Length 1705;
  Best Local Similarity   90.6%;  Pred. No. 2e-226;
  Matches  978;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60

Db     62 ATGACATCAACATCCAAAGGCATTCTTCGTCCATTCCTAATCGTCTGTGTTATCCTGGCC 121

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120

Db    122 TGCTTCATGGCATGTCTGCTTATTTACATCAAGCCCACCAACAGCTGGGTGTTCAGTCCA 181

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180

Db    182 ATGGAGTCTGCCAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 241

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240

Db    242 TTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 301

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300

Db    302 TCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACAGACCGCTCACTGTAC 361

Qy    301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360

Db    362 AACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACTTA 421

Qy    361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420

Db    422 CCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAACCTAGAGTCTCCCACT 481

Qy    421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480

Db    482 CATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACTCTAACATATCGCCGT 541

Qy    481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540

Db    542 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 601

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600

Db    602 GAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAACTGGAATCCGGAGCAT 661

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660

Db    662 GCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATCCATACCTATGGGCAA 721

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720

Db    722 GCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATATCTACTTGTAAATTT 781

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780

Db    782 TATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAAAAGCTCTACAATGCA 841

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840

Db    842 TTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAAAACTATGAGAATTAT 901

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900

Db    902 ATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGTTGGCAAAA 961

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960

Db    962 TATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTATTTCAACTGGAGGAAG 1021

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
```

```
Db   1022 GATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGCCTGGCTTGTGATCAC 1081

Qy   1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
Db   1082 GTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAATGGTTTTGGAATTAA 1141


RESULT 8
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match             72.6%;  Score 783.6;  DB 5;  Length 1707;
  Best Local Similarity   82.8%;  Pred. No. 5.5e-192;
  Matches  894;  Conservative    1;  Mismatches  185;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
Db     268 ATGACATCAACATCTAAAGGAATTTTCCGGCCATTTTTTATTGTCTTCATTGTYCTTGGT 327

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
Db     328 TGTTTCATGGCATTTATACTAATTTATATTAAACCAACAAGCAGCTGGATCTCTAATCCT 387

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
Db     388 GTGGAATCAGCCAGCTCAGTTTTGAAAATGAAGAGCTTCTTTTCTTCCAAAACTGATAAT 447

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240

Db     448 TTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGACATTCGATCTAACA 507

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
Db     508 TCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTGACCGCTCACTATAT 567

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
Db     568 AACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGGATCTGACTAATTTA 627

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
Db     628 CCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACTTGGAATCTCCAACT 687

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
Db     688 CATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCCTGACTTACCGGCGT 747

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
Db     748 GATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAGGCGCCTTTACATAT 807

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
Db     808 GAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACTGGAACCCTGAGCAT 867

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
Db     868 GCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCCATACCTACGGGCAA 927

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
Db     928 GCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCTCCAGTTGCAAATTC 987

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
Db     988 TACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGAAACTTTACAATGCT 1047

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
Db    1048 CTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAAACTACGAGAATTAC 1107

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
Db    1108 ATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAAGAGAGCTTGCAGAA 1167

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
Db    1168 TATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACTTCAACTGGAGGAAG 1227

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
Db    1228 GATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTCTTGCTTGTGATCAT 1287

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
Db    1288 GTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAATGGTTTTGGAACTAA 1347


RESULT 9
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
```

```
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN

  Query Match             62.7%;  Score 676.8;  DB 5;  Length 1080;
  Best Local Similarity   76.7%;  Pred. No. 2.3e-164;
  Matches  828;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
           ||||||||| |||| ||||||||  | ||  ||||||||  |||| ||  | |  ||  |
Db       1 ATGACATCAGCATCTAAAGGAATCTTACGGGCATTTTTAGCTGTCAGCGGTGTGTTGATC 60

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
           ||||| |||  |||| | || || ||| ||||||| || |||| |||||| | ||||||
Db      61 TGTTTTATGATATGTTTACTGATATACGTCAAACCAACAAACAACTGGATTTCCAGTCCT 120

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
           || ||||||||    || || || |||||||| || || || ||| | |||    ||||||
Db     121 ATCGAATCAGCACATTCAGTCCTTAAAATGAAGAATTTTTTCTCCTCTAAACAGAATTAT 180

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
           |  ||||||||     ||||  | ||||||||||| |||||||||| ||||| || | |
Db     181 TACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGACTTTTGAACTGAAA 240

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
           ||||||||  || |||| || ||  | || |||||||| || || |||||  |||| |||
Db     241 TCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTGACCGCGCACTCTAC 300

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
           ||||| || ||||| || ||  | || ||| ||||||| || |||||||| |  ||  |
Db     301 AACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGGATCTCATTAACCTG 360

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
           || || ||    |||||||| |||||||||||||||||||||||| ||||||||||||
Db     361 CCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATCTGGAATCACCAACA 420

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
           || || || ||||| |||||||||||||||||| |||||||||||||||||||| | |||
Db     421 CATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTCTGACTTACAGGCGT 480

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
           || ||||||||||||||||||||||||||| ||   |||||||||| || || |  |||
Db     481 GACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAAAACCATTTGACTTT 540

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
           || |||||||||||||| || || || || |||||||| || |||||||||| || |||
Db     541 GAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACTGGAACCCAGAACAT 600

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
           |||||||||||| |||||||||| || | |||| |||||||||    || || ||||||
Db     601 GCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTATAACTTATGGGCAA 660

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
           || ||||| |||||  | |  |||||||  || |||| || ||  |  |||| ||||||
Db     661 GCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTGCATCTTGCAAATTT 720

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
           ||||| || |||||||||||| || ||||| ||||| ||||| ||||| || || |||||
Db     721 TATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAAAACTTTATAATGCA 780

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
           ||  ||||||| || || ||  ||||  |||||||  |||| |||||||||||||| ||
Db     781 TTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAAACTATGAGAACTAC 840

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
           ||||| |||||||| || || || |||||||||| |  |||||| || ||  ||||  |
Db     841 ATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAAGAGAATTAGCTGAC 900

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
           ||||        |  ||||| | |   |     ||||| || ||||  ||||||||||
Db     901 CATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACTACAACTGGAGGAAA 960

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGCATGTTTGGCTTGCGATCAT 1020
           ||||  || || ||| | || |  || |||||||| |||||||| || || || || |||
Db     961 CATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCTTAGCGTGTGACCAT 1020

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
           |||||||||||||||||||| || || || || ||| ||||||| ||||| |||| |||
Db    1021 GTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGTGGTTCTGGAACTAA 1080
```

```
RESULT 10
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```

```
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1702: contig of 814 bp in length
            *     1703 1802: gap of      100 bp
            *     1803     2642: contig of 840 bp in length
            *     2643 2742: gap of      100 bp
            *     2743     3637: contig of 895 bp in length
            *     3638 3737: gap of      100 bp
            *     3738     4635: contig of 898 bp in length
            *     4636 4735: gap of      100 bp
            *     4736     5657: contig of 922 bp in length
            *     5658 5757: gap of      100 bp
            *     5758     6692: contig of 935 bp in length
            *     6693 6792: gap of      100 bp
            *     6793     7703: contig of 911 bp in length
            *     7704 7803: gap of      100 bp
            *     7804     8742: contig of 939 bp in length
            *     8743 8842: gap of      100 bp
            *     8843     9787: contig of 945 bp in length
            *     9788 9887: gap of      100 bp
            *     9888    10819: contig of 932 bp in length
            *    10820 10919: gap of      100 bp
            *    10920    11808: contig of 889 bp in length
            *    11809 11908: gap of      100 bp
            *    11909    12830: contig of 922 bp in length
            *    12831 12930: gap of      100 bp
            *    12931    13847: contig of 917 bp in length
            *    13848 13947: gap of      100 bp
            *    13948    14886: contig of 939 bp in length
            *    14887 14986: gap of      100 bp
            *    14987    15911: contig of 925 bp in length
            *    15912 16011: gap of      100 bp
            *    16012    16903: contig of 892 bp in length
            *    16904 17003: gap of      100 bp
            *    17004    17906: contig of 903 bp in length
            *    17907 18006: gap of      100 bp
            *    18007    18931: contig of 925 bp in length
            *    18932 19031: gap of      100 bp
            *    19032    19945: contig of 914 bp in length
            *    19946 20045: gap of      100 bp
            *    20046    20951: contig of 906 bp in length
            *    20952 21051: gap of      100 bp
            *    21052    21949: contig of 898 bp in length
            *    21950 22049: gap of      100 bp
            *    22050    22950: contig of 901 bp in length
            *    22951 23050: gap of      100 bp
            *    23051    23945: contig of 895 bp in length
            *    23946 24045: gap of      100 bp
            *    24046    24955: contig of 910 bp in length
            *    24956 25055: gap of      100 bp
            *    25056    25980: contig of 925 bp in length
            *    25981 26080: gap of      100 bp
            *    26081    27052: contig of 972 bp in length
            *    27053 27152: gap of      100 bp
            *    27153    28067: contig of 915 bp in length
            *    28068 28167: gap of      100 bp
            *    28168    29076: contig of 909 bp in length
            *    29077 29176: gap of      100 bp
            *    29177    30093: contig of 917 bp in length
            *    30094 30193: gap of      100 bp
            *    30194    31099: contig of 906 bp in length
            *    31100 31199: gap of      100 bp
            *    31200    32097: contig of 898 bp in length
            *    32098 32197: gap of      100 bp
            *    32198    33113: contig of 916 bp in length
            *    33114 33213: gap of      100 bp
            *    33214    34125: contig of 912 bp in length
            *    34126 34225: gap of      100 bp
            *    34226    35133: contig of 908 bp in length
            *    35134 35233: gap of      100 bp
            *    35234    36150: contig of 917 bp in length
            *    36151 36250: gap of      100 bp
            *    36251    37171: contig of 921 bp in length
            *    37172 37271: gap of      100 bp
            *    37272    38159: contig of 888 bp in length
            *    38160 38259: gap of      100 bp
            *    38260    39156: contig of 897 bp in length
            *    39157 39256: gap of      100 bp
            *    39257    40188: contig of 932 bp in length
            *    40189 40288: gap of      100 bp
            *    40289    41210: contig of 922 bp in length
            *    41211 41310: gap of      100 bp
            *    41311    42222: contig of 912 bp in length
            *    42223 42322: gap of      100 bp
            *    42323    43218: contig of 896 bp in length
            *    43219 43318: gap of      100 bp
            *    43319    44233: contig of 915 bp in length
            *    44234 44333: gap of      100 bp
            *    44334    45246: contig of 913 bp in length
            *    45247 45346: gap of      100 bp
            *    45347    46261: contig of 915 bp in length
            *    46262 46361: gap of      100 bp
            *    46362    47279: contig of 918 bp in length
            *    47280 47379: gap of      100 bp
            *    47380    48276: contig of 897 bp in length
            *    48277 48376: gap of      100 bp
            *    48377    49282: contig of 906 bp in length
            *    49283 49382: gap of      100 bp
            *    49383    50305: contig of 923 bp in length
            *    50306 50405: gap of      100 bp
            *    50406    51314: contig of 909 bp in length
            *    51315 51414: gap of      100 bp
            *    51415    52335: contig of 921 bp in length
            *    52336 52435: gap of      100 bp
            *    52436    53163: contig of 728 bp in length
            *    53164 53263: gap of      100 bp
            *    53264    54210: contig of 947 bp in length
            *    54211 54310: gap of      100 bp
            *    54311    55199: contig of 889 bp in length
            *    55200 55299: gap of      100 bp
            *    55300    56212: contig of 913 bp in length
            *    56213 56312: gap of      100 bp
            *    56313    57287: contig of 975 bp in length
            *    57288 57387: gap of      100 bp
            *    57388    58245: contig of 858 bp in length
            *    58246 58345: gap of      100 bp
            *    58346    59236: contig of 891 bp in length
            *    59237 59336: gap of      100 bp
            *    59337    60272: contig of 936 bp in length
```

```
            *    60273 60372: gap of      100 bp
            *    60373    61289: contig of 917 bp in length
            *    61290 61389: gap of      100 bp
            *    61390    62299: contig of 910 bp in length
            *    62300 62399: gap of      100 bp
            *    62400    63322: contig of 923 bp in length
            *    63323 63422: gap of      100 bp
            *    63423    64336: contig of 914 bp in length
            *    64337 64436: gap of      100 bp
            *    64437    65354: contig of 918 bp in length
            *    65355 65454: gap of      100 bp
            *    65455    66352: contig of 898 bp in length
            *    66353 66452: gap of      100 bp
            *    66453    67362: contig of 910 bp in length
            *    67363 67462: gap of      100 bp
            *    67463    68370: contig of 908 bp in length
            *    68371 68470: gap of      100 bp
            *    68471    69379: contig of 909 bp in length
            *    69380 69479: gap of      100 bp

  Query Match             59.2%;  Score 639;  DB 2;  Length 71396;
  Best Local Similarity   85.0%;  Pred. No. 1.2e-154;
  Matches  764;  Conservative    0;  Mismatches  122;  Indels   13;  Gaps    4;

Qy      72 ATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGC 131
           || |||    | | |      |  || ||||||||||| | ||||||||||||||||| || ||
Db   65456 ATCTCTCGCCCTGTGGTGGATTCCCACCAACAGCTCGGTCTTCAGTCCAATGGAGTCTGC 65515

Qy     132 CAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAAC 191
           || ||||||||||||||||||| ||||| ||||| |||||||||||||||||| |||||
Db   65516 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 65575

Qy     192 TACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 251
           ||| |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db   65576 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 65635

Qy     252 AATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCA 311
           |||||||| |||||||| |||||||||||||| ||||| ||| ||||||||||||||||
Db   65636 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 65695

Qy     312 TGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGC 371
           ||| || ||||| || ||  |||||||||| |||||||||| || ||||||||||| ||
Db   65696 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 65755

Qy     372 TAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCA 431
           |||||||||||| |||||||||||||||||||||||| || ||||| |||||||| |||||
Db   65756 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 65815

Qy     432 AAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATAT 491
           |||||||||||||||| |||||||| |||||||||||| ||||| ||||||||||||||||
Db   65816 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 65875

Qy     492 CCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAG 551
           |||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||
Db   65876 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 65935

Qy     552 CAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAA 611
           ||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65936 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 65995

Qy     612 GTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGA 671
           ||||||||| ||||| ||||| || |||||||||| ||||| || |||||||| |||||
Db   65996 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 66055

Qy     672 ATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTT 731
           ||| || || |||||||||| ||||||| |||||||| |||||||||||||||||||  |
Db   66056 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCTCT 66115

Qy     732 TGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGG 791
             ||| |      || |  | | |   | |         | ||     ||| |||||||
Db   66116 GAAAACTCATTCACAAGATTACTCCCAAAAG---------TCCACATGCTTTTTGGCTGG 66166

Qy     792 CTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGA 851
```

```
           || |||||||||| ||||| ||||||||||||||||||||||||||||||||||||| ||
Db   66167 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 66226

Qy     852 TTCATTCATTC--ATGTGGAAGATTATAACTCT-CCCAGTGAGCTAGCAAAGTATCTGAA 908
           ||||||||||    |||||||||| | ||||| |||||||| |||||||    |||||
Db   66227 TTCATTCATTCCATCGTGGAAGATTTTCACTCTCCCCAGTGAGTTAGCAAATATCCTGAA 66286

Qy     909 GGAAGTCGACAAAAACAATAAGTT-ATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 966
           |||||||||||| ||||||||||  |||||||||||| ||||| |  ||| |  |||
Db   66287 GGAAGTCGACAANAACAATAAGTTGGTACCTTAGTTACTCTAACTCGNAGAAAGGATTC 66345


RESULT 11
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN

  Query Match             25.7%;  Score 277.8;  DB 5;  Length 1567;
  Best Local Similarity   58.5%;  Pred. No. 3.3e-61;
  Matches  527;  Conservative    0;  Mismatches  362;  Indels   12;  Gaps    2;

Qy     183 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 242
           | | ||||| |   |  || | |||||||  || || ||       |||||| |   |
Db     461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy     243 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 302
           ||     |  ||||||||||||| || || ||| | |||   ||| |     ||  | |
Db     521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy     303 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 362
```

```
Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGG---------ATTTGCC 631

Qy    363 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 422

Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy    423 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 482

Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy    483 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTT 539

Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy    540 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 599

Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy    600 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 659

Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy    660 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 719

Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy    720 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 779

Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy    780 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 839

Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy    840 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 899

Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy    900 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 959

Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy    960 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 1019

Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy   1020 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1079

Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy   1080 A 1080
          |
Db   1352 A 1352
```

```
RESULT 12
AB023628
LOCUS       AB023628               1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN
```

```
  Query Match           21.7%;  Score 234.6;  DB 5;  Length 1832;
  Best Local Similarity 55.5%;  Pred. No. 4.9e-50;
  Matches  500;  Conservative    0;  Mismatches  389;  Indels   12;  Gaps    2;

Qy    183 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 242

Db    619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy    243 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 302

Db    679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy    303 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 362

Db    739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGC---------ACAGATTTTCC 789

Qy    363 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 422

Db    790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy    423 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 482

Db    850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy    483 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA---AATCCCTTCGTGTT 539

Db    910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy    540 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 599

Db    970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy    600 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 659

Db   1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy    660 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 719

Db   1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy    720 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 779
```

```
Db   1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy    780 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 839

Db   1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy    840 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 899

Db   1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy    900 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 959

Db   1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy    960 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 1019

Db   1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy   1020 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1079

Db   1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy   1080 A 1080

Db   1510 A 1510
```

```
RESULT 13
AB035907
LOCUS       AB035907                1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN
```

```
  Query Match           20.6%;  Score 222.6;  DB 5;  Length 1092;
  Best Local Similarity   57.3%;  Pred. No. 6.3e-47;
  Matches  484;  Conservative    0;  Mismatches  349;  Indels   12;  Gaps    4;

Qy    184 AATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 243

Db    199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy    244 TGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAAC 303

Db    259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy    304 AAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCT 363

Db    319 GAGGCAGACGCCATTATGTTCCACCACCGGGACAT---TTCTGATTTCAGTGACCTTCCT 375

Qy    364 CAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 423

Db    376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy    424 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 483

Db    436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy    484 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 543

Db    496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy    544 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 603

Db    553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy    604 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 663

Db    613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy    664 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 723

Db    670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy    724 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 780

Db    730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840

Db    790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900

Db    850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960

Db    910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020

Db    970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy   1021 GTGAA 1025

Db   1030 GTCAA 1034
```

```
RESULT 14
AF288369
LOCUS       AF288369                1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
```

```
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN

  Query Match           20.4%;  Score 220.2;  DB 5;  Length 1264;
  Best Local Similarity   58.0%;  Pred. No. 2.6e-46;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    6;  Gaps    3;

Qy    193 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 252
Db    175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy    253 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
Db    235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy    313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 372
Db    295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy    373 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 432
Db    355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy    433 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 492
Db    415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy    493 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 552
Db    475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy    553 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 612
Db    535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy    613 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 672
Db    595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGC-GGCAACACTTGCC 653

Qy    673 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 732
Db    654 CCTGTCCATG--GACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy    733 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATAC---AATGCTTTTCTGGCT 789
Db    712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy    790 GGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCA 849
Db    772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy    850 GATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAG 909
Db    832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy    910 GAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
Db    892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942


RESULT 15
GGU73678
LOCUS       GGU73678               2657 bp    DNA     linear   VRT 20-DEC-1996
DEFINITION  Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete
            cds.
ACCESSION   U73678
VERSION     U73678.1  GI:1657998
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N.,
            Schultz,T.L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
  TITLE     Molecular cloning and characterization of CFT1, a developmentally
            regulated avian alpha(1,3)-fucosyltransferase gene
  JOURNAL   J. Biol. Chem. 271 (51), 32960-32967 (1996)
  MEDLINE   97115837
REFERENCE   2  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M. and Larsen,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical
            Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES             Location/Qualifiers
     source          1. .2657
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            170. .1240
                     /gene="CFT1"
     CDS             170. .1240
                     /gene="CFT1"
                     /note="similar to mammalian fucosyltransferase FucTIV"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase"
                     /protein_id="AAC60060.1"
                     /db_xref="GI:1657999"
                     /translation="MELGPRWSPAARPGCPRRWRRRWALLGALLGAALALYVCVRELR
                     RRGSAAGRPEGEVTVLLWWEPFGRPWRPADCRRRYNITGCLLSADRGRYGEARAVLFH
                     HRDLALHGRQGLPRGPPPRPPRQRWVWMNFESPSHSPGLRGLAGLFNWTMSYRRDSDV
                     FVPYGYLYEPPSPRPFVLPRKSRLVAWVISNWNEEHARVRYYRQLKEHLPIDVYGARG
                     MALLEGSVVKTVSAYKFYLAFENSQHTDYITEKLWKNAFAASAVPVVLGPRRANYERF
                     IPADSFIHVDDFPSPRLLATYLKFLDKNKPSYRRYFAWRNKYEVHVTSFWDEHYCKVC
                     EAVRTAGNQLKTVQNLAGWFES"
BASE COUNT      581 a    712 c    818 g    546 t
ORIGIN

  Query Match           16.3%;  Score 176.2;  DB 5;  Length 2657;
  Best Local Similarity   53.1%;  Pred. No. 6.1e-35;
  Matches  451;  Conservative    0;  Mismatches  383;  Indels   15;  Gaps    3;

Qy    193 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 252
Db    341 ACGGTGCTGCTGTGGTGGGAGCCCTTCGGCCGCCCCTGGCGCCCGGCCGACTGCCGGAGG 400
```

```
Qy    253 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
Db    401 CGCTACAACATCACGGGCTGCCTCCTGAGCGCCGACCGCGGGCGGTACGGCGAGGCTCGG 460

Qy    313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAG-- 370
Db    461 GCCGTGCTCTTCCACCACCGCGACCTGGCGCTGCACGGCCGCCAGGGGCTGCCCCGCGGG 520

Qy    371 -------CTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 423
Db    521 CCCCCGCCGCGACCCCCGCGGCAGCGGTGGGTGTGGATGAACTTCGAGTCGCCCTCGCAC 580

Qy    424 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 483
Db    581 TCCCCCGGGCTGCGGGGCCTGGCCGGGCTCTTCAACTGGACCATGTCGTACCGACGGGAC 640

Qy    484 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 543
Db    641 TCGGACGTCTTCGTGCCCTACGGGTACCTCTACGAGCCGCCGTCGCCGCGGCCCTTCGTG 700

Qy    544 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 603
Db    701 CTGCCTCGCAAAAGCCGGCTGGTGGCCTGGGTCATCAGCAACTGGAACGAGGAGCACGCC 760

Qy    604 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 663
Db    761 CGCGTGCGCTACTACCGGCAGCTGAAGGAGCACCTGCCCATCGACGTGTACGGG---GCG 817

Qy    664 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 723
Db    818 CGCGGGATGGCGCTGCTGGAGGGCAGCGTGGTGAAGACGGTGTCGGCCTACAAGTTCTAC 877

Qy    724 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 780
Db    878 CTGGCCTTCGAGAACTCCCAGCACACGGACTACATCACCGAGAAGCTGTGGAAGAACGCC 937

Qy    781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
Db    938 TTCGCCGCCAGCGCGGTGCCTGTTGTCCTCGGCCCCCGCAGGGCCAACTACGAGCGCTTC 997

Qy    841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
Db    998 ATCCCCGCCGACTCCTTCATCCACGTGGACGACTTCCCCAGCCCACGGCTGCTGGCCACC 1057

Qy    901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
Db   1058 TACCTGAAGTTCCTCGATAAAAACAAGCCCAGCTACAGGAGGTATTTCGCCTGGCGGAAC 1117

Qy    961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
Db   1118 AAGTATGAAGTCCACGTCACGTCTTTCTGGGATGAGCATTACTGCAAGGTTTGCGAGGCC 1177

Qy   1021 GTGAAAAGG 1029
Db   1178 GTGAGGACG 1186
```

Search completed: October 6, 2002, 04:35:48
Job time : 1497.36 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 121.198 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-4_COPY_1289_2368
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1080	100.0	2676	21	AAZ92647	Human alpha-1,3-fu
	2	1080	100.0	2822	21	AAZ92646	Human alpha-1,3-fu
	3	931.2	86.2	2036	21	AAZ92672	Murine alpha-1,3-f
	4	931.2	86.2	2170	21	AAZ92645	Murine alpha-1,3-f
	5	127.4	11.8	1814	18	AAT59505	Murine myeloid-lin
	6	127	11.8	3594	18	AAT85219	Mouse alpha-fucosy
	7	114.4	10.6	1701	15	AAQ77732	Human alpha-1,3-fu
	8	114.4	10.6	1701	21	AAF21136	Human low adenosin
	9	114.4	10.6	1701	21	AAA35014	Human adenosine re
	10	114.4	10.6	6944	21	AAF21137	Human low adenosin
	11	114.4	10.6	6944	21	AAA35015	Human adenosine re
	12	113.4	10.5	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
	13	113.4	10.5	1654	18	AAT61680	Human alpha(1,3)-f
	14	112.2	10.4	1086	15	AAQ56912	Lewis blood group
	15	112.2	10.4	2042	18	AAT80111	Fucosyltransferase
	16	112.2	10.4	2043	12	AAQ13330	Glycosyltransferas
	17	112.2	10.4	2043	15	AAQ56906	DNA encoding a gly
	18	112.2	10.4	2043	18	AAT76769	Human alpha 1,3/4
	19	112.2	10.4	2043	18	AAT61675	Human alpha(1,3/1,
	20	111.8	10.4	1126	21	AAF21135	Human low adenosin
	21	111.8	10.4	1126	21	AAA35013	Human adenosine re
	22	111.8	10.4	1316	15	AAQ56911	DNA encoding a gly
	23	111.8	10.4	1316	18	AAT61679	Human alpha(1,3)-f
	24	103.6	9.6	795	17	AAT13798	Alpha-(1,3/1/4)-fu
	25	102.4	9.5	2175	11	AAQ06691	CDX, a MILA involv
	26	100.8	9.3	1256	21	AAF21134	Human low adenosin
	27	100.8	9.3	1256	21	AAA35012	Human adenosine re
	28	100.8	9.3	1400	13	AAQ31436	Encodes a HeLa cel
	29	100.8	9.3	1488	15	AAQ56910	DNA encoding a gly
	30	100.8	9.3	2175	12	AAQ14382	Clone 7.2 encoding
	31	100.8	9.3	2175	24	AAS17082	Human cDNA clone 7
	32	100.8	9.3	2861	12	AAQ14383	Clone 1 encoding 1
	33	100.8	9.3	2861	21	AAF21133	Human low adenosin
	34	100.8	9.3	2861	21	AAA35011	Human adenosine re
	35	100.8	9.3	2861	24	AAS17083	Human cDNA clone 1
	36	100.8	9.3	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
	37	100.8	9.3	3647	15	AAQ56909	DNA encoding a gly
	38	100.8	9.3	3647	18	AAT61678	Human alpha(1,3)-f
	39	97.6	9.0	2134	18	AAT59506	Human myeloid deri
	40	66.6	6.2	1578	23	ABL11867	Drosophila melanog
c	41	66.6	6.2	4229	23	ABL11866	Drosophila melanog
	42	52.2	4.8	1353	21	AAZ51685	C. elegans alpha-1
	43	44.4	4.1	676	22	AAS41344	cDNA encoding nove
	44	44.4	4.1	1479	24	ABA05334	Human fucosyltrans
	45	44.4	4.1	2557	24	ABA05333	Human fucosyltrans

ALIGNMENTS

RESULT 1
AAZ92647
ID AAZ92647 standard; cDNA; 2676 BP.
XX
AC AAZ92647;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 295..1374
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

```
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match           100.0%;  Score 1080;  DB 21;  Length 2676;
  Best Local Similarity  100.0%;  Pred. No. 4e-300;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 654

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 714

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 774

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 1074

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1134

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1194

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1195 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1255 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1314

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1374


RESULT 2
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag= a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
```

```
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 134-142; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

  Query Match             100.0%;  Score 1080;  DB 21;  Length 2822;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-300;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1289 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 1348

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1349 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 1408

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1409 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 1468

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1469 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 1528

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1529 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 1588

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1589 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 1648

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1649 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 1708

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1709 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 1768

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1769 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 1828

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1829 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 1888

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1889 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 1948

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1949 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 2008

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2009 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 2068

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2069 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 2128

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2129 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 2188

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2189 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 2248

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2249 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 2308

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2309 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 2368


RESULT 3
AAZ92672
ID   AAZ92672 standard; cDNA; 2036 BP.
XX
AC   AAZ92672;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
```

```
XX
PI   Narimatsu H, Kudo T, Sasaki K;
XX
DR   WPI; 2000-183120/16.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer. -
XX
PS   Example 8; Page 162-164; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

  Query Match             86.2%;  Score 931.2;  DB 21;  Length 2036;
  Best Local Similarity   91.4%;  Pred. No. 2.2e-257;
  Matches  987;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db       9 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 68

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
           || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db      69 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 128

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTTCCACCAAAACTGATTAT 180
           ||||| || || || ||||||||||||||||||||| ||||| ||||| ||||||||||||
Db     129 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 188

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
           ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db     189 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 248

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
           |||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db     249 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 308

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
           |||||||||||||| || ||||| || ||  |||||||||| |||||||||| || |||
Db     309 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 368

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
           |||||||| || |||||||||||| |||||||||||||||||||||||| || ||||| |||
Db     369 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 428

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
           ||||| |||||||||||||||||||| |||||||| |||||||||| ||||| ||||||
Db     429 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 488

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db     489 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 548

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
           |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db     549 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 608

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
           ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db     609 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 668

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
           ||||| |||||||| || || ||||||||| ||||||| ||||||||| ||||||||||
Db     669 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 728

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
           |||||||| |||||||| ||||| ||||| |||||||||| |||||||| ||||||||
Db     729 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 788

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
           ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||||||||||
Db     789 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 848

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
           |||||||| |||||||||||||||||||||||||| |||||||||||||| |||||||
Db     849 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 908

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
           |||||||||||||| |||||||||||||||| |||||||||||||||||||||||| |||
Db     909 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 968

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
           ||||| |||||||| || ||||| |||||||||||||||||||  |||| ||||| |||
Db     969 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1028

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1029 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1088

RESULT 4
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag= a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
```

```
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match             86.2%;  Score 931.2;  DB 21;  Length 2170;
  Best Local Similarity   91.4%;  Pred. No. 2.2e-257;
  Matches  987;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db     115 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 174

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
           || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db     175 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 234

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
           ||||| || || || |||||||||||||||||| ||||| ||||| ||||||||||||
Db     235 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 294

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
           ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||
Db     295 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 354

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
           |||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db     355 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 414

Qy     301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 360
           |||||||||||||| || ||||| || ||  |||||||||| |||||||||| || |||

Db     415 AACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTA 474

Qy     361 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 420
           |||||||| || |||||||||| |||||||||||||||||||||| || ||||| |||
Db     475 CCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACT 534

Qy     421 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 480
           ||||| |||||||||||||||||||| |||||||| |||||||||| ||||| ||||||
Db     535 CACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGT 594

Qy     481 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db     595 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTT 654

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
           |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db     655 GAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 714

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
           ||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| || |||
Db     715 GCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAA 774

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
           ||||| |||||||| || || ||||||||| ||||||| |||||||| ||||||||||
Db     775 GCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTT 834

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 780
           |||||||| |||||||| ||||| ||||| |||||||||| |||||||| ||||||||
Db     835 TATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCA 894

Qy     781 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 840
           ||| ||||||| || |||||||||| ||||| ||||||||||||||||||||||||||
Db     895 TTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTAT 954

Qy     841 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 900
           |||||||| |||||||||||||||||||||| |||||||||||||| |||||||
Db     955 ATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAA 1014

Qy     901 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
           |||||||||||||| |||||||||||||||| |||||||||||||||||||||| |||
Db    1015 TATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAG 1074

Qy     961 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1020
           ||||| |||||||| || ||||| ||||||||||||||||||  |||| ||||| |||
Db    1075 GATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCAT 1134

Qy    1021 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1080
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 GTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag= a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
```

```
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             11.8%;  Score 127.4;  DB 18;  Length 1814;
  Best Local Similarity   52.1%;  Pred. No. 3e-26;
  Matches  334;  Conservative    0;  Mismatches  301;  Indels    6;  Gaps    2;

Qy     269 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 328
           | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     329 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 388
           |||| ||  |          |   |     |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     389 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 448
           |||  | |||       |||||||  ||  |  |  | ||  |          | ||  |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     449 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           |  |  ||  |||  |   ||||  ||  ||  ||||||||||||||||   ||  ||  ||  ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     509 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 568
           ||||  |     |  |    ||   |    |     | |  ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 628
           ||||||  |  ||  ||  |    |       |||      |  |  |||   |||     ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     629 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 688
           |       |  |  |  ||   | ||| |  ||    |||        |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     689 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 748
           | ||  |  ||  ||   |  |     |      ||  ||  ||   |||||||  ||  |||    ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     749 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 805
           |||  ||||||||  ||  |||  |  |    ||||||   |   ||||||   |||||||  ||  |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     806 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 865
           ||||||||  ||  |||  |  |||  |||   ||  |  |  |||  |||||  |  ||   |  || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     866 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 906
           ||||  ||  |  |  ||||  ||  ||||  ||  ||     |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag= a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag= b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag= c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag= d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag= e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag= f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag= g
FT                   /number= 3
FT   exon            2067..3552
FT                   /*tag= f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
```

```
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             11.8%;  Score 127;  DB 18;  Length 3594;
  Best Local Similarity   52.3%;  Pred. No. 5.2e-26;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     269 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 328
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336

Qy     329 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 388
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     389 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 448
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     449 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     509 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 568
Db    2517 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 628
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     629 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 688
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     689 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 748
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy     749 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 805
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy     806 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 865
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy     866 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904
```

```
RESULT 7
AAQ77732
ID   AAQ77732 standard; cDNA to mRNA; 1701 BP.
XX
AC   AAQ77732;
XX
DT   22-JUN-1995  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; sialylated Lewis blood group antigen;
KW   Lewis x; Lewis a; glycosylation; glycoprotein; glycolipid; ds.
XX
OS   Homo sapiens (THP-1 cells).
XX
FH   Key             Location/Qualifiers
FT   CDS             123..1151
FT                   /*tag= a
FT                   /product= alpha-1,3-fucosyltransferase
XX
PN   WO9423021-A.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994;   94WO-JP00496.
XX
PR   29-MAR-1993;   93JP-0069016.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hanai N,  Kurata K,  Nishi T,  Sasaki K;
XX
DR   WPI; 1994-333183/41.
DR   P-PSDB; AAR63215.
XX
PT   Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
PT   of active sugar chains, e.g. sialylated Lewis X
XX
PS   Claim 3; Page 99-102; 155pp; Japanese.
XX
CC   The sequence coding for alpha-1,3-fucosyltransferase was isolated
CC   from a human THP-1 cell cDNA library. The enzyme catalyses the
CC   addition of fucose via a 1,3-linkage to N-acetylglucosamine
CC   contained in glycoproteins and glycolipids, e.g. for sialylation of
CC   Lewis blood group antigens.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             10.6%;  Score 114.4;  DB 15;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 1.6e-22;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     256 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 315
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     316 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 375
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     376 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 435
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     436 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 495
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     496 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 555
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629
```

```
Qy    556 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 615
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    616 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 675
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    676 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 735
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    736 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 792
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    793 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 852
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    853 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 8
AAF21136
ID AAF21136 standard; DNA; 1701 BP.
XX
AC AAF21136;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2703.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 1701 BP; 290 A; 527 C; 589 G; 295 T; 0 other;

Query Match 10.6%; Score 114.4; DB 21; Length 1701;
Best Local Similarity 50.8%; Pred. No. 1.6e-22;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy    256 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 315
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    316 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 375
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    376 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 435
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    436 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 495
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    496 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 555
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    556 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 615
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    616 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 675
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    676 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 735
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    736 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 792
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    793 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 852
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    853 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
```

```
                | ||| | |||||||| || | |  |||  || | ||||| ||
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 9
AAA35014
ID AAA35014 standard; DNA; 1701 BP.
XX
AC AAA35014;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2703.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 957-958; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

```
  Query Match             10.6%;  Score 114.4;  DB 21;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 1.6e-22;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy    256 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 315
          | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    316 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 375
          ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    376 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 435
          || |     |||  ||| | |||       |||| || || |  ||||| | |
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    436 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 495
          ||   | |      | | || |||  |  ||||  ||||| || || || || |||
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    496 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 555
          |||||  ||||||  | ||  |     |       ||||    |       |||||  |||
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    556 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 615
            |  |||    |||||| || || ||||   |      |||     | | || |
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    616 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 675
          |||    ||||       |   |  ||  ||  || |  ||   ||||
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    676 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 735
          |    |  |      || | || ||| |  |    |     || || || ||||||||
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    736 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 792
          || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    793 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 852
          |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    853 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
          | ||| | |||||||| || | |  |||  || | ||||| ||
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 10
AAF21137
ID AAF21137 standard; DNA; 6944 BP.
XX
AC AAF21137;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2704.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX

PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034-1036; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

Query Match 10.6%; Score 114.4; DB 21; Length 6944;
Best Local Similarity 50.8%; Pred. No. 3e-22;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy     256 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 315
           | |  ||||    | |||||| || |   |  ||||    |||    ||   || | ||
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     316 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 375
           ||  |  |||| ||||| ||  |          | | |  |  | || | |      ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     376 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 435
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     436 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 495
           ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     496 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 555
           ||||| ||||||  | ||  |      |       ||||    |       ||||| |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     556 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 615
           |   |||    |||||| || || ||||   |       |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     616 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 675
           |||    ||||           |      |  ||  ||  || | |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     676 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 735
           |   |  |      || | || ||| |  |     |      || || || ||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy     736 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 792
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy     793 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 852
           |||| || || || ||||| ||  |  |||  | ||||||   |   | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy     853 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

RESULT 11
AAA35015
ID AAA35015 standard; DNA; 6944 BP.
XX
AC AAA35015;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2704.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 958-960; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which

```
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

  Query Match            10.6%;  Score 114.4;  DB 21;  Length 6944;
  Best Local Similarity  50.8%;  Pred. No. 3e-22;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     256 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 315
           | |  ||||    | |||||  || |  |  ||||    |||     ||   || | ||
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     316 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 375
           ||  |  |||| |||||  ||  |          | |  |  |  | || | |       ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     376 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 435
           || |     |||   ||| | |||       |||| || || |  |||||  | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     436 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 495
           ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     496 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 555
           ||||| ||||||  | ||  |      |      ||||    |       |||||  |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     556 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 615
              |  |||    |||||| || || ||||   |      |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     616 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 675
           |||    ||||         |    |  ||  ||  || |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     676 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 735
           |    |  |       || | || ||| |  |    |      || || || ||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy     736 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 792
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy     793 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 852
           ||||  || || || |||||  ||  |  |||  | |||||||    |   | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy     853 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

```
RESULT 12
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag= a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase. The enzyme produced by the
CC   DNA may be non glycosylated. This prevents premature loss of enzyme
CC   activity. It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match            10.5%;  Score 113.4;  DB 15;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 3e-22;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
           || ||| |||||  ||||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
           |       ||      |  |||  |  ||||  | |||||      ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429
           | ||||  |      |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
```

```
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540

Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600

Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660

Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720

Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777

Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837

Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897

Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957

Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     958 AAG 960

Db    1095 GAG 1097
```

RESULT 13
AAT61680
ID AAT61680 standard; DNA; 1654 BP.
XX
AC AAT61680;
XX
DT 19-JUN-1997 (first entry)
XX
DE Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
XX
KW Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW glycosylation; oligosaccharide; blood group; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 129..1208
FT /*tag= a
XX
PN WO9709421-A1.
XX
PD 13-MAR-1997.
XX
PF 06-SEP-1996; 96WO-US13816.
XX
PR 08-SEP-1995; 95US-0525058.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Legault DJ, Lowe JB;
XX
DR WPI; 1997-192897/17.
DR P-PSDB; AAW13643.
XX
PT New recombinant fucosyltransferase proteins - useful for modifying
PT cell surface oligosaccharide structures
XX
PS Example 6; Page 291-292; 329pp; English.
XX
CC A DNA clone (AAT61680) codes for an enzyme (AAW13643) that functions as
CC a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC precursors but not type I glycoconjugates. It was identified in a
CC human lambda phage genomic DNA library using a probe obtd. from
CC Lewis enzyme cDNA (see also AAT61675). The Fuc-TVI DNA can be used
CC to construct animal cell lines with specific capabilities with
CC respect to post-translational modification of the oligosaccharides
CC of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC use in oligosaccharide prodn., to genotype individuals at this
CC fucosyltransferase locus, and in the prodn. of chimeric
CC fucosyltransferases (see also AAW14512-32).
XX
SQ Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

Query Match 10.5%; Score 113.4; DB 18; Length 1654;
Best Local Similarity 51.2%; Pred. No. 3e-22;
Matches 401; Conservative 0; Mismatches 361; Indels 21; Gaps 5;

```
Qy     196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255

Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312

Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369

Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429

Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489

Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540

Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600

Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660

Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720

Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777

Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837

Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897

Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957
```

```
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    958 AAG 960
Db   1095 GAG 1097


RESULT 14
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
DE   Lewis blood group fucosyltransferase DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding a Lewis blood group fucosyl
CC   transferase (Fuc-TIII).  The enzyme produced by the DNA may be
CC   non glycosylated.  This prevents premature loss of enzyme activity.
CC   It can also be used in in vitro reactions to modify cell surface
CC   oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-11.
XX
SQ   Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

  Query Match            10.4%;  Score 112.2;  DB 15;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 5.5e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 15
AAT80111
ID   AAT80111 standard; cDNA; 2042 BP.
XX
AC   AAT80111;
XX
DT   01-DEC-1997  (first entry)
XX
DE   Fucosyltransferase III coding sequence.
XX
KW   P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW   calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW   thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW   inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW   disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW   adult respiratory distress syndrome; tumour metastasis;
KW   fucosyltransferase III; acetylglucosaminyltransferase;
KW   autoimmune disorder; inflammatory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9706176-A2.
XX
PD   20-FEB-1997.
XX
PF   02-AUG-1996;   96WO-US12820.
XX
PR   17-MAY-1996;   96US-0649802.
PR   03-AUG-1995;   95US-0510920.
PR   15-MAY-1996;   96US-0017794.
XX
PA   (OKLA ) UNIV OKLAHOMA STATE.
XX
PI   Cummings RD,  Mcever RP,  Moore KL;
XX
DR   WPI; 1997-154206/14.
XX
PT   New O-glycan(s) derived from P-selectin glyco:protein ligand
PT   (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT   for treating inflammatory or autoimmune disorders or tumours
XX
PS   Disclosure; Page 80-81; 96pp; English.
XX
```

CC This sequence represents the coding sequence for the human
CC fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC protein expressed on the surfaces of activated platelets and endothelium
CC in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC high affinity P-selectin ligand produced by leukocytes. Binding of
CC P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC highly extended extracellular domain, which is a feature of mucin-like
CC proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC O-glycans, and O-glycanated products of the invention. The products are
CC used for inhibiting the binding of P-selectin and other selectins to
CC PSGL-1. They can be used for the treatment of inflammation, injury
CC resulting from ischaemia and reperfusion, bacterial sepsis and
CC disseminated intravascular coagulation, adult respiratory distress
CC syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC other autoimmune or inflammatory disorders. The products can also be used
CC in the production of antibodies and in diagnostic applications.
XX
SQ Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

Query Match 10.4%; Score 112.2; DB 18; Length 2042;
Best Local Similarity 51.6%; Pred. No. 7.3e-22;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
          |   ||  |    |  |    |     |           | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
           |    ||            |  |  |  |    |||||  |   |                |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
          || |  |  |   ||||    ||  | ||     |  || || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
          || || || || ||||| ||  .  ||    ||||| ||||  |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
          |||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
```

```
          |||||  || |       |||||  || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 5, 2002, 23:14:29
Job time : 158.198 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 904.255 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-4_COPY_1289_2368
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	573.8	53.1	697	9	BB397507	BB397507 BB397507
2	492.8	45.6	607	10	BI102760	BI102760 602888338
3	422	39.1	716	9	AV725071	AV725071 AV725071
4	324.8	30.1	649	9	AV338075	AV338075 AV338075
c 5	307.4	28.5	579	10	BM021081	BM021081 ie75f03.x
c 6	296.8	27.5	1101	12	CNS059ST	AL327638 Tetraodon
7	284.6	26.4	648	9	BB595164	BB595164 BB595164
8	272.4	25.2	625	9	BB649967	BB649967 BB649967
9	261.6	24.2	767	9	BB640911	BB640911 BB640911
10	243	22.5	568	10	BM021317	BM021317 ie75f03.y
11	219.6	20.3	628	9	BB651409	BB651409 BB651409
c 12	184.2	17.1	570	10	BM082484	BM082484 fu25e08.x
c 13	177	16.4	1100	12	CNS0579Z	AL324368 Tetraodon
c 14	170.2	15.8	1016	12	CNS0429P	AL271222 Tetraodon
15	165	15.3	844	12	CNS01V8V	AL168808 Tetraodon
c 16	157.4	14.6	887	12	CNS02Q0S	AL208693 Tetraodon
17	155.4	14.4	651	10	BJ036855	BJ036855 BJ036855
c 18	151.6	14.0	666	10	BJ050300	BJ050300 BJ050300
c 19	151.2	14.0	711	12	CNS034XF	AL228012 Tetraodon
c 20	151	14.0	501	12	B41073	B41073 HS-1052-B2-
21	146.6	13.6	640	10	BJ073227	BJ073227 BJ073227
22	140.6	13.0	533	10	BI604066	BI604066 603243886
23	136.6	12.6	670	9	BB626126	BB626126 BB626126
24	136.2	12.6	199	9	BB398103	BB398103 BB398103
c 25	123.4	11.4	674	10	BE605503	BE605503 fl10h04.x
26	120.6	11.2	971	12	CNS047IV	AL278032 Tetraodon
c 27	120.4	11.1	615	10	BI326811	BI326811 STR00078
c 28	117.8	10.9	1023	12	CNS02ANG	AL188773 Tetraodon
29	102.8	9.5	487	10	BE863396	BE863396 UI-M-BH0-
c 30	100.2	9.3	702	10	BJ043578	BJ043578 BJ043578
31	93.8	8.7	863	12	CNS02ROF	AL210840 Tetraodon
32	90.4	8.4	720	10	BF180380	BF180380 601804564
33	90	8.3	551	10	BJ035701	BJ035701 BJ035701
34	87.4	8.1	601	10	BJ036679	BJ036679 BJ036679
35	87.4	8.1	605	10	BJ028804	BJ028804 BJ028804
c 36	87	8.1	1056	12	CNS01VL7	AL169252 Tetraodon
37	86.8	8.0	304	9	AU081478	AU081478 AU081478
38	84.2	7.8	638	10	BJ035697	BJ035697 BJ035697
39	83.6	7.7	1057	12	CNS03ADC	AL235065 Tetraodon
c 40	83	7.7	566	10	BJ090565	BJ090565 BJ090565
c 41	79.4	7.4	776	12	CNS03RIO	AL257289 Tetraodon
42	78	7.2	838	10	BG432556	BG432556 602500710
43	76.4	7.1	655	10	BG538740	BG538740 602566977
c 44	75	6.9	878	12	CNS02V1D	AL215194 Tetraodon
45	71.6	6.6	590	10	BM082755	BM082755 fu25e08.y

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN
```

```
  Query Match             53.1%;  Score 573.8;  DB 9;  Length 697;
  Best Local Similarity   91.4%;  Pred. No. 7e-145;
  Matches  619;  Conservative    0;  Mismatches   57;  Indels    1;  Gaps    1;

Qy     404 ATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGA 463
           | || || ||||| |||||||| ||| |||||||||||| ||| |||||||| |||||||
Db       4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy     464 CTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCA 523
           |||| ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||
Db      63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy     524 CAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTA 583
           |||||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||
Db     123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy     584 ACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAA 643
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||
Db     183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy     644 TCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCA 703
           |||| ||||| || |||||||| |||||||| || || |||||||| ||||||| ||||
Db     243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy     704 TATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGG 763
```

```
           ||||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |
Db     303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy     764 AAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGG 823
           ||||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db     363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy     824 AAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTC 883
           |||||||||||||||||||||||| |||||||||||||||||||||| ||||||||
Db     423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy     884 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTT 943
           |||||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db     483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy     944 ACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCAT 1003
           |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||
Db     543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy    1004 GTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 1063
           |  |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db     603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy    1064 AATGGTTTTGGAATTAA 1080
           |||||||||||||||||
Db     663 AATGGTTTTGGAATTAA 679
```

```
RESULT 2
BI102760
LOCUS       BI102760                607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: l  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN
```

```
  Query Match             45.6%;  Score 492.8;  DB 10;  Length 607;
  Best Local Similarity   90.1%;  Pred. No. 6.5e-123;
  Matches  549;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;
```

```
Qy    223 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACA 282
          |||||||||||| |||||||||||||||||||||||||| |||||||| ||||||||||||
Db      1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy    283 ACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGT 342
          || ||||| ||| |||||||||||||||||| || ||||| || ||  ||||||||||
Db     61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy    343 TGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATG 402
          |||||||||| || |||||||||| || |||||||||| |||||||||| |||||||
Db    120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy    403 AATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTG 462
          ||||| || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||
Db    179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy    463 ACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGC 522
          ||||| ||||| |||||||||||||||||||||||||||||||||||||||||||| |||
Db    239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy    523 ACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGT 582
          |||||||||| |||||||||||||||||| ||||| ||||||||||||||||||||||
Db    299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy    583 AACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAA 642
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||
Db    359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy    643 ATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACC 702
          ||||| ||||| || |||||||| |||||||| || || |||||||||| ||||||| |||
Db    419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy    703 ATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACG 762
          |||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||
Db    479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy    763 GAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGG 822
          |||||||| |||||||| ||| ||||||| || ||||||  ||| ||||| |||||||||
Db    539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy    823 GAAAACTAT 831
          |||||||||
Db    599 GAAAACTAT 607
```

RESULT 3
AV725071
LOCUS AV725071 716 bp mRNA linear EST 16-OCT-2000
DEFINITION AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION AV725071
VERSION AV725071.1 GI:10830091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 716)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA HTC clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .716
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBLB02"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 219 a 113 c 133 g 249 t 2 others
ORIGIN

Query Match 39.1%; Score 422; DB 9; Length 716;
Best Local Similarity 100.0%; Pred. No. 1.1e-103;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    659 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy    719 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy    779 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy    839 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy    899 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy    959 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 1018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy   1019 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy   1079 AA 1080
          ||
Db    421 AA 422
```

RESULT 4
AV338075
LOCUS AV338075 649 bp mRNA linear EST 24-OCT-2001
DEFINITION AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION AV338075
VERSION AV338075.2 GI:16396626
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 649)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)

```
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430407H24"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory bulb"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      209 a    105 c    117 g    218 t
ORIGIN

  Query Match             30.1%;  Score 324.8;  DB 9;  Length 649;
  Best Local Similarity   91.5%;  Pred. No. 2.7e-77;
  Matches  344;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy     705 ATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGA 764
           ||| ||||||| ||| |||||| |||||||| ||||| ||||||||||||||||| ||
Db       2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61

Qy     765 AAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGA 824
           |||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db      62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy     825 AAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCC 884
           ||||||||||||||||||||||||| ||||||||||||||||||||||| |||||||||
Db     122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy     885 CAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTA 944
           ||||||| ||||||| |||||||||||||| ||||||||||||||||| ||||||||||
Db     182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy     945 CTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATG 1004
           |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db     242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy    1005 TTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 1064
             |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db     302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy    1065 ATGGTTTTGGAATTAA 1080
           ||||||||||||||||
Db     362 ATGGTTTTGGAATTAA 377


RESULT 5
BM021081/c
LOCUS       BM021081                579 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021081
VERSION     BM021081.1  GI:16535437
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
```

/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 203 a 103 c 93 g 180 t
ORIGIN

Query Match 28.5%; Score 307.4; DB 10; Length 579;
Best Local Similarity 99.7%; Pred. No. 1.4e-72;
Matches 308; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    772 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy    832 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy    892 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy    952 TGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCT 1011
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy   1012 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 1071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy   1072 TGGAATTAA 1080
          |||||||||
Db    279 TGGAATTAA 271
```

RESULT 6
CNS059ST/c
LOCUS CNS059ST 1101 bp DNA linear GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 045P22 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL327638
VERSION AL327638.1 GI:8221227
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="045P22"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA045DH11A1~end : T3"
BASE COUNT 287 a 249 c 269 g 278 t 18 others
ORIGIN

Query Match 27.5%; Score 296.8; DB 12; Length 1101;
Best Local Similarity 58.3%; Pred. No. 1.3e-69;
Matches 603; Conservative 0; Mismatches 423; Indels 9; Gaps 5;

```
Qy     50 TTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGA 109
          | ||| |||| || |   || |         | || |||   || ||  ||| ||||| |
Db   1064 TCATCGTGGGGTGCTNTGTGACTGTGTNNTTTATGTACTNTAATCCATCCACCAGCT-GN 1006

Qy    110 TCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCA 169
          | | | ||||  ||||  |||          |  | |  | || ||     | || ||||
Db   1005 TTTCCGGTCCTGTGGAGGCAGGACTGAGCTCGGAGCATGTCAAGAATCAACTATCAACCA 946

Qy    170 AAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCT 229
          | | |||| |    ||    || | ||| ||| ||||| | ||||| || || || ||||
Db    945 AGAGTGATAAA---AACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCT 889

Qy    230 TTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACC 289
          | ||||| |     ||| | ||    |||  | |  | || ||  |  | ||  |||||
Db    888 TCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACA 829

Qy    290 GTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATC 349
          |    || |  ||||| ||  |||  ||| |  |||| ||| ||||||||    | || |
Db    828 GAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACC 769

Qy    350 TGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGG 409
          || | ||  | ||      |   | ||  || ||||||||||||| ||| |  |  |||
Db    768 TGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGG 709

Qy    410 AATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGA 469
          | || ||  | ||| |  |      |  ||| || |||  | |||| |||||| ||||
Db    708 AGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGA 649

Qy    470 CTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATC 529
            || || |  ||||| || ||         ||  |   || |      ||   ||| ||
Db    648 ACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATG 589

Qy    530 CCTTCGTG---TTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 586
          |    | |   ||||      || |||||| | ||| |||| || ||| |||| || ||||
Db    588 CTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACT 529

Qy    587 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 646
          ||||||  || |||||  |||| || |  ||||| ||  |   |||      |||| |||
Db    528 GGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCC 469

Qy    647 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 706
          | || || || | ||||||  ||  || |  ||   ||| || |     | || |||||
Db    468 ACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCG 409

Qy    707 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 766
          |    || || || || || || ||||| || || |||||||| || ||| | || ||||
```

```
Db    408 CCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAA 349

Qy    767 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 826
          | || |||||  |  ||   |  ||  | || || || |||||||| ||| || |||| |
Db    348 AACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGA 289

Qy    827 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 886
          |||| |||||    || |  | |||| | || || || |||||||| | |  ||| |||
Db    288 ACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCA 229

Qy    887 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 946
            |||||  ||  | ||| |   |    | ||||||||  |  |  | |||||||  || |
Db    228 AGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATT 169

Qy    947 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 1006
          ||  |   | ||||   |      ||       |  ||||      |   | |||
Db    168 TTCANGTGGCGAAGACACTTTAAAGCAAGGGCAGTGGCATTTGCCCGAGGC-CACCTGTC 110

Qy   1007 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAA 1065
           |||||| ||| | || |  |||| | |  |||| | ||| |   |  |  ||| | ||
Db    109 AGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAG 50

Qy   1066 TGGTTTTGGAATTAA 1080
          ||||  |||  || |
Db     49 TGGTACTGGGGTTGA 35


RESULT 7
BB595164
LOCUS       BB595164                 648 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB595164 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION   BB595164
VERSION     BB595164.2  GI:16450167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11491766.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C030004N01"
                     /clone_lib="RIKEN full-length enriched, adult male corpus
                     striatum"
                     /sex="male"
                     /tissue_type="corpus striatum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      141 a    206 c    138 g    163 t
ORIGIN

  Query Match             26.4%;  Score 284.6;  DB 9;  Length 648;
  Best Local Similarity   90.0%;  Pred. No. 2.2e-66;
  Matches  305;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          |||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db    310 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 369

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    370 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 429

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db    430 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 489

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| |||| |||||||||||||||||||||||||||||||
Db    490 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 549

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          |||||||||||||||||||| |||| ||| |||||||||||||| ||||| ||| |||||
Db    550 TCCTGCCAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 609

Qy    301 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATC 339
```

```
          ||||||||||||||| || ||||| || ||   |||||||
Db    610 AACAAATCCCATGCGGTCCTGATTCACCATATAGACATC 648


RESULT 8
BB649967
LOCUS       BB649967                625 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB649967 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION   BB649967
VERSION     BB649967.1  GI:16484222
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C130099N17"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="16 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT      136 a    198 c    135 g    156 t
ORIGIN

  Query Match             25.2%;  Score 272.4;  DB 9;  Length 625;
  Best Local Similarity   91.7%;  Pred. No. 4.4e-63;
  Matches  288;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          |||||||||||||||||||| |||||||||||||||| |||| |||||||| |||||||||
Db    311 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 370

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    371 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 430

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||| || || || ||||||||||||||||||| ||||| ||||| ||||||||||||
Db    431 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 490

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db    491 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 550

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
          |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db    551 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 610

Qy    301 AACAAATCCCATGC 314
          ||||||||||||||
Db    611 AACAAATCCCATGC 624


RESULT 9
BB640911
LOCUS       BB640911                767 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB640911 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730093G08 5', mRNA sequence.
ACCESSION   BB640911
VERSION     BB640911.1  GI:16476069
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
```

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A730093G08"
                     /clone_lib="RIKEN full-length enriched, 7 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="7 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      175 a    239 c    176 g    172 t      5 others
ORIGIN

  Query Match             24.2%;  Score 261.6;  DB 9;  Length 767;
  Best Local Similarity   88.0%;  Pred. No. 4.1e-60;
  Matches  285;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
           ||||||||||||||||||||| |||||||||||||||| |||| |||||||| |||||||||
Db     347 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 406

Qy      61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
           || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db     407 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 466

Qy     121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
           ||||| || || || |||| |||||||||||||| ||||| ||||| ||||||||||||
Db     467 ATGGAGTCTGCAAGTTCTGGGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 526

Qy     181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
           ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db     527 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 586

Qy     241 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 300
           ||| |||||||||||||||| ||||| ||  ||||||| |||||  | || ||| || ||
Db     587 TCCGGCCAAGCAATGTTCAATATCCAGGGGGGCCATCTTACAACAAAACGGTCATTGGAC 646

Qy     301 AACAAATCCCATGCAGTTCTGATC 324
           ||||||||||| |  |  ||||||
Db     647 AACAAATCCCAGGGGGGCCTGATC 670
```

RESULT 10
BM021317

```
LOCUS       BM021317                568 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021317
VERSION     BM021317.1  GI:16535673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ie75f03.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
```

by hydroxyapatite chromatography and used to make this library."
BASE COUNT 127 a 170 c 123 g 148 t
ORIGIN

Query Match 22.5%; Score 243; DB 10; Length 568;
Best Local Similarity 100.0%; Pred. No. 4.1e-55;
Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 385

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    386 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 445

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    446 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 505

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    506 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 565

Qy    241 TCC 243
          |||
Db    566 TCC 568
```

RESULT 11
BB651409
LOCUS BB651409 628 bp mRNA linear EST 26-OCT-2001
DEFINITION BB651409 RIKEN full-length enriched, 0 day neonate cerebellum Mus musculus cDNA clone C230079N02 5', mRNA sequence.
ACCESSION BB651409
VERSION BB651409.1 GI:16485662
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 628)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .628
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230079N02"
/clone_lib="RIKEN full-length enriched, 0 day neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 129 a 200 c 143 g 152 t 4 others
ORIGIN

Query Match 20.3%; Score 219.6; DB 9; Length 628;
Best Local Similarity 88.1%; Pred. No. 9.6e-49;
Matches 237; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy      1 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 60
          |||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db    347 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 406

Qy     61 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 120
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    407 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 466

Qy    121 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 180
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db    467 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 526

Qy    181 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 240
          |||| |||||||| |||||||  | ||| ||||||| |||||||||| ||||||||||
Db    527 GTTAACGAAACTACCATTCTGGNTTAGGTATGGCCATNTGGGCAGACCTNTGACCTTACA 586

Qy    241 TCCTGCCAAGCAATGTTCAACATCCAAGG 269
          |||||||||   |||||  | ||||||||
Db    587 TCCTGCCAAAGCATGTTTCATATCCAAGG 615
```

RESULT 12
BM082484/c
LOCUS BM082484 570 bp mRNA linear EST 14-NOV-2001
DEFINITION fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223 3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;,

```
            mRNA sequence.
ACCESSION   BM082484
VERSION     BM082484.1  GI:16929414
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN
```

```
  Query Match             17.1%;  Score 184.2;  DB 10;  Length 570;
  Best Local Similarity   59.2%;  Pred. No. 3.8e-39;
  Matches  337;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    1;

Qy     382 TTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGC 441
           |||||||| |||||||| |  |||| ||||| || ||  |||||      ||    |||
Db     570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy     442 ATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCT 501
           | ||  || |||| ||||  ||||| |  ||| | |  ||| | ||||| | ||| |
Db     510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy     502 TATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAA 561
             | ||    ||     |  | |    ||    |  ||||    | || |  || || |||
Db     450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy     562 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGAGTCAAG 612
             | || || ||| ||||||||||||     |||          ||      || ||   || |
Db     390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331

Qy     613 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 672
             |||    ||  |||||  |  |||  |  | ||    | |||| | || | |  |  |
Db     330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271
```

```
Qy     673 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 732
           ||| |  |  ||||| | |    |||||| |  ||   || |||||||| || |||||
Db     270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211

Qy     733 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 792
           || || || |||||||||||| |||||||| || ||| |  |   | || ||   || ||
Db     210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151

Qy     793 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 852
           |||| ||||| ||| |||| ||  | ||| ||||||||||| || |  |||||  | |||
Db     150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91

Qy     853 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 912
           | || |||||||| | || |       |  |||    ||||||   ||| || |
Db      90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31

Qy     913 GTCGACAAAAACAATAAGTTATACCTTAG 941
           |  | |||||  | |    ||| |  ||
Db      30 CTAAATAAAAATGACATTGCATATCACAG 2
```

```
RESULT 13
CNS0579Z/c
LOCUS       CNS0579Z                1100 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            007H05 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL324368
VERSION     AL324368.1  GI:9557250
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007H05"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA007CD03C1-end : T7"
BASE COUNT      314 a    228 c    252 g    304 t      2 others
ORIGIN
```

```
  Query Match             16.4%;  Score 177;  DB 12;  Length 1100;
  Best Local Similarity   55.4%;  Pred. No. 4.2e-37;
  Matches  428;  Conservative    0;  Mismatches  332;  Indels   13;  Gaps    4;
```

```
Qy    196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
          || ||| | ||| | ||||| | ||| |||| |||| |  ||  ||  |||
Db    773 ATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTTGCATACGACTCTTGTAAAGTTTAC 714

Qy    256 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 315
          | ||| ||  | || ||  | ||||||   || || || || ||| ||||| | || ||
Db    713 TACAATATTGACGGCTGTGAGCTCACAGATGATCGATCCCTTTACGACAAAGCGCAGGCT 654

Qy    316 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 375
          ||||||  ||   |||   |||||    ||| |||||  ||| || ||   | |  |  ||
Db    653 GTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGGAAACTTGCCAGTGGAGCCGCGG 594

Qy    376 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 435
          |||   || ||||  ||||||||| |  |  ||||||| || |   ||||   |   |
Db    593 CCATATTTTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGATA 534

Qy    436 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 495
            ||| | ||     | || ||| |||||||||  ||| |    ||||| || ||   |
Db    533 CCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGTA 474

Qy    496 GTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 552
          |    ||   |     | | |  |     | |    |      | |  ||||
Db    473 GCCAGGTACCCTCTTACCATCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGAG 414

Qy    553 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC------ATGCCAGA 606
          ||| | ||| | |  || ||| | || || ||    |   |            | | |||
Db    413 AAAAACAAACTTGCCTGTTGGATCGTTAGCAATGCTACATCCACTGGGACAGGTACAAGA 354

Qy    607 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 666
          |   | | ||| ||||| |||  ||||   ||| | ||   |   |  ||    |  |
Db    353 GCGCAATTTTATAATGAACTATCCAAACATATTAACATTAGTGTTTTTGGTGTGGTGTAC 294

Qy    667 GGAGAATATGTCAATGATAAAAATTTGATT---CCTACCATATCTGCTTGTAAATTTTAT 723
            ||          || |   ||  |   |  || ||| |  ||   ||||| |||||
Db    293 ACAGGGGTCAGGTTGGAGATTGATCAGTATTACCCGACCGTTGCTAGCTGTAAGTTTTAC 234

Qy    724 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTT 783
          | || ||||| || ||| | ||||| ||||||||||| || ||| |  ||    |  ||
Db    233 TTATCATTTGAGAACTCATTACACAAAGATTACATCACAGAGAAGGTCAACGNACCCCTT 174

Qy    784 CTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATT 843
            || ||  | || ||||| ||||||||||  | || |  ||||| || |  |   |
Db    173 GCAGCAGGAACCGTTCCTGTAGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTTC 114

Qy    844 CCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTAT 903
          ||  |||| || || ||||| || || |||| |     ||| |  |  || ||| || ||
Db    113 CCTTCAGACTCCTTTATTCACGTAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCAT 54

Qy    904 CTGAAGGAAGTCGACAA-AAACAATAAGTTATACCTTAGTTACTTTAACTGGA 955
          |||       | |||||   |  |  ||  |||| |  | |||||| | ||||
Db     53 CTGCTCTCCTTGGACAAGGNATGAGGAGGCATACATGCGCTACTTTGAGTGGA 1
```

```
RESULT 14
CNS0429P/c
LOCUS       CNS0429P                1016 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            075P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271222
VERSION     AL271222.1  GI:7993186
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1016)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1016
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="075P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG075DH01SP1-end :
                     PUC-Ori"
BASE COUNT      246 a    216 c    253 g    298 t      3 others
ORIGIN
```

```
  Query Match           15.8%;  Score 170.2;  DB 12;  Length 1016;
  Best Local Similarity   57.0%;  Pred. No. 2.9e-35;
  Matches  394;  Conservative    0;  Mismatches  283;  Indels   14;  Gaps    4;

Qy    197 TTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGT 256
          ||||| ||||| | |||||  | ||   |  |||||  | |     |||    |   |
Db    715 TTCTGCTGTGGTTCTGGCCCATCGGCACGGTGTTTGATTTCAAGGATTGCTCCACCCACT 656

Qy    257 TCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAG 316
          |||| ||  |  | || |||||  ||     ||| | || |||||||  |||| |  | | ||
Db    655 TCAAGATTGACAGCTG-CATCTGACGGACGACAGATCTCTGTACAGCAAAGCAAAGGGAG 597

Qy    317 TTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGC 376
          |  |  || | |||  |||||  |  |||  |        |  || ||   |    |  | |
Db    596 TCATTTTCTACCACAAAGACACAACATGGCACACAGACAGCTTGCCCACGGGTCCACGCC 537

Qy    377 CACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGA 436
          | |  |||||||  ||| |||||  | |||  |||||||  |||||   ||||    |   ||
Db    536 CCCATTTCCAGAGGTGGGTTTGGTTTAATGTGGAATCCCCAACAAACACATACAGGAAAC 477

Qy    437 GTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAG 496
            ||  | ||  || |||| |||||||| ||||  ||| |  | ||    |  ||    |
Db    476 CAGGTCTGGAAAACCTGTTCAACCTGACGCTGAGCTACAGGAGAGACGGCGGCATTGTGG 417

Qy    497 TGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTG---TTTGAAGTGCCAAGCA 553
            |    | |   | |   |    | | ||  |  | ||  ||||   |||| |   |
Db    416 CGAGGAACGAGATGTCTATCCGAGACCCAGGTGACACCGAGGACTTTGTTCTGCCCAAGA 357

Qy    554 AAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATG---------CCA 604
          ||||  ||   |||||| ||| | || || |||    ||||        |         |||
Db    356 AAGACAAGACGGTGTGTTGGATCGTCAGCAACAATGACCCGAGAACAGGAACAGGCACCA 297

Qy    605 GAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCAT 664
          ||| | | || |||  ||| ||   ||||  |||  |  |  |    |  ||  |||| |
Db    296 GAGACCAATACTACCGTGACCTGTCCAAACACATCAAGGTGGACGTTTGGGGCAAAGCCT 237

Qy    665 TTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATC 724
          ||||  | |   | | |  | |  | |    ||| || ||  |    |||||||| ||||
Db    236 TTGGGAATTTCCTGAGTTTTGAGGAATACTATCCAACAATTGCCAGCTGTAAATTCTATC 177

Qy    725 TTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTC 784
          | || || || || || ||||||| || ||||| ||||| ||| |  ||    | |||
Db    176 TGTCTTTCGAGAACTCCATCCACAAAGACTACATGACGGAGAAGATCAACGGGCCCTTTG 117
```

```
Qy    785 TGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATAT-T 843
Db    116 TGGCTGGGTCAGTTCCTATAGTTTTGGGACCTCCAAAAAAAAAAAAAAAAAAAATTCTAT 57

Qy    844 CCAGCAGATTCATTCATTCATGTGGAAGATT 874
Db     56 CCTTCCGATTCGTTTATTCACGTTAACGACT 26


RESULT 15
CNS01V8V
LOCUS       CNS01V8V                 844 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            198K09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL168808
VERSION     AL168808.1  GI:7806865
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="198K09"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG198AF05SP1~end :
                     PUC-Ori"
BASE COUNT      199 a    190 c    205 g    245 t      5 others
ORIGIN

  Query Match             15.3%;  Score 165;  DB 12;  Length 844;
  Best Local Similarity   54.1%;  Pred. No. 7e-34;
  Matches  399;  Conservative    3;  Mismatches  324;  Indels   12;  Gaps    3;

Qy     20 GAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTC 79
Db    119 RAACCATGCGGCCCTTTTTCATTGCTGTTTTTGTTTTGGGGGCCTGTGGAACTTTAATCT 178

Qy     80 TCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTG 139
Db    179 TTATGCTTTGGAAACCATCCACTCGCTGGCTGTTAGGTCCTGTGGAAACA------ACAG 232

Qy    140 TGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTC 199
Db    233 CAGCACCTGTYAGAGAAATCATTTCCCTCAGGACGGATAA---AAACCTGACCACGGTGC 289

Qy    200 TGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCA 259
Db    290 TGCTCTGGTTCTGGCCATTTGGGATTACCTTTGACCCGAACGTGTGCAGCGCTTACTATG 349

Qy    260 ACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTC 319
Db    350 GTATAGACGGCTGTTTCATTACAGCGGACAGGAACTATTTCAGCGAAGGAATGCGGNTCG 409

Qy    320 TGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCAC 379
Db    410 TTATTCATCATCGAGAMATTTCTCCCGACCTGTCCAATCTGCCGCAGCAGGAGCGGCCGT 469

Qy    380 CCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTG 439
Db    470 TGTACCARATGTGGGTCTGGTTCAACATGGAGAATCCTTCGTACTCCGACAGGTTTCTTG 529

Qy    440 GCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGC 499
Db    530 GAATGGACGGCTTGTTCAACCTGACTCTGAATTATCGTCTGGATGCTGATATTTCCATGC 589

Qy    500 CTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGA 559
Db    590 CTCATGTGTACCTGGTTCCTGAAAAAAGTGAGGAAGACTTTGTCCCCCCCAGCAAAGATA 649

Qy    560 AATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACA 619
Db    650 AGTTGGTCTGCTGGATAGTCAGCAACTGGAAGGAGCATTATATACGGGTGAAATACTTCA 709

Qy    620 ATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCA 679
Db    710 AGGAGCTGTCCAAGCACATAAAGATCACCACTTATGGACGTGGCTTTGGGGGCTA---CA 766

Qy    680 ATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATT 739
Db    767 ACTGGGGAAGTGCCTCGTCCATCATGGCCAGCTGTAAGTTCTACCTGTCTTTTGAGAACT 826

Qy    740 CAATCCACAAGGATTACA 757
Db    827 CCATCTACATTGACTACA 844

Search completed: October  6, 2002, 07:16:06
Job time : 912.255 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 22.0361 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-4_COPY_1289_2368
Perfect score: 1080
Sequence: 1 atgacatcaacatccaaagg.....agaaatggttttggaattaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	11.8	1814	2	US-08-483-151-1	Sequence 1, Appli
2	127	11.8	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	113.4	10.5	1654	1	US-07-914-281-13	Sequence 13, Appl
4	113.4	10.5	1654	1	US-08-393-246-13	Sequence 13, Appl
5	113.4	10.5	1654	1	US-08-525-058A-13	Sequence 13, Appl
6	113.4	10.5	1654	2	US-08-696-731-13	Sequence 13, Appl
7	113.4	10.5	1654	4	US-09-042-531-13	Sequence 13, Appl
8	112.2	10.4	1086	1	US-07-914-281-12	Sequence 12, Appl
9	112.2	10.4	1086	1	US-08-393-246-12	Sequence 12, Appl
10	112.2	10.4	1086	1	US-08-525-058A-12	Sequence 12, Appl
11	112.2	10.4	1086	2	US-08-696-731-12	Sequence 12, Appl
12	112.2	10.4	1086	4	US-09-042-531-12	Sequence 12, Appl
13	112.2	10.4	2042	3	US-09-063-237-2	Sequence 2, Appli
14	112.2	10.4	2043	1	US-07-914-281-1	Sequence 1, Appli
15	112.2	10.4	2043	1	US-08-393-246-1	Sequence 1, Appli
16	112.2	10.4	2043	1	US-08-273-411-4	Sequence 4, Appli
17	112.2	10.4	2043	1	US-08-525-058A-1	Sequence 1, Appli
18	112.2	10.4	2043	2	US-08-696-731-1	Sequence 1, Appli
19	112.2	10.4	2043	4	US-09-042-531-1	Sequence 1, Appli
20	112.2	10.4	2043	5	PCT-US91-00899-1	Sequence 1, Appli
21	111.8	10.4	1316	1	US-07-914-281-10	Sequence 10, Appl
22	111.8	10.4	1316	1	US-08-393-246-10	Sequence 10, Appl
23	111.8	10.4	1316	1	US-08-525-058A-10	Sequence 10, Appl
24	111.8	10.4	1316	2	US-08-696-731-10	Sequence 10, Appl
25	111.8	10.4	1316	4	US-09-042-531-10	Sequence 10, Appl
26	100.8	9.3	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	100.8	9.3	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	100.8	9.3	1488	1	US-07-914-281-9	Sequence 9, Appli
29	100.8	9.3	1488	1	US-08-393-246-9	Sequence 9, Appli
30	100.8	9.3	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	100.8	9.3	1488	2	US-08-696-731-9	Sequence 9, Appli
32	100.8	9.3	1488	4	US-09-042-531-9	Sequence 9, Appli
33	100.8	9.3	2175	4	US-08-482-073-9	Sequence 9, Appli
34	100.8	9.3	2861	4	US-08-482-073-10	Sequence 10, Appl
35	100.8	9.3	3647	1	US-07-914-281-7	Sequence 7, Appli
36	100.8	9.3	3647	1	US-08-393-246-7	Sequence 7, Appli
37	100.8	9.3	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	100.8	9.3	3647	2	US-08-696-731-7	Sequence 7, Appli
39	100.8	9.3	3647	4	US-09-042-531-7	Sequence 7, Appli
40	100.8	9.3	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	97.6	9.0	2134	2	US-08-483-151-3	Sequence 3, Appli
42	97.6	9.0	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	37.2	3.4	1728	1	US-08-403-866-12	Sequence 12, Appl
c 44	37.2	3.4	2231	1	US-08-403-866-14	Sequence 14, Appl
c 45	37.2	3.4	12720	1	US-08-403-866-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 11.8%; Score 127; DB 2; Length 1814;
Best Local Similarity 52.3%; Pred. No. 3e-28;
Matches 330; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

```
Qy    269 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 328
          | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy     329 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 388
           |||| ||  |           |   |     |||| | | |    |||||||  |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     389 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 448
             ||| | |||       |||||||  || | | | || |           | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     449 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           | | || |||  |   ||||  || || ||||||||||||||    || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     509 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 568
           ||||  |     | |    ||   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 628
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     629 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 688
           |       | |  |  ||   | ||| |  ||    |||      |  | |||  ||||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     689 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 748
           | || | || ||  |  |    ||||||  |   ||||||  |||||||  || |  | |
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     749 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 805
           ||| ||||||||| || ||| | |    ||||||  |   ||||||  |||||||  || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     806 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 865
           ||||||||  || |||  | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     866 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
           |||| || |  | |||| || |||| || ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06427
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US96-06427-1

Query Match 11.8%; Score 127; DB 5; Length 1814;
Best Local Similarity 52.3%; Pred. No. 3e-28;
Matches 330; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

```
Qy     269 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 328
           | |||| ||| |   |  ||||   |||     ||   |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     329 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 388
           |||| ||  |           |   |     |||| | | |    |||||||  |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     389 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 448
             ||| | |||       |||||||  || | | | || |           | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     449 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 508
           | | || |||  |   ||||  || || ||||||||||||||    || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     509 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 568
           ||||  |     | |    ||   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     569 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 628
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     629 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 688
           |       | |  |  ||   | ||| |  ||    |||      |  | |||  ||||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     689 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 748
           | || | || ||  |  |    ||||||  |   ||||||  |||||||  || |  | |
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     749 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 805
           ||| ||||||||| || ||| | |    ||||||  |   ||||||  |||||||  || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     806 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 865
           ||||||||  || |||  | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     866 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 896
           |||| || |  | |||| || |||| || ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

RESULT 3
US-07-914-281-13
; Sequence 13, Application US/07914281
; Patent No. 5324663
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-13

  Query Match            10.5%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 3.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
          || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
          |      ||      |  ||| |  ||||  | |||||     ||||      |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489
          ||   ||  | ||| ||     |  || || || ||  || | ||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||     ||| || ||||          | | | |     |   | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |  |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          |||||  ||    ||  |||  |    ||           |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
            |      |     |     |   ||   ||||    ||  | ||     |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777
          |||||  |||| || || ||  | |||   || ||||||||  || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837
          ||  |   |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897
          |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957
            ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    958 AAG 960
            ||
Db   1095 GAG 1097


RESULT 4
US-08-393-246-13
; Sequence 13, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-13

  Query Match            10.5%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 3.2e-24;
```

```
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
          || ||| ||||| |||||| ||| |  | || | |  ||  |  |||| || |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
           |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429
          | ||||  |     |||  ||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489
          ||   ||  | ||| ||     |  || || || ||  || | ||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||     ||| || ||||           | | | |      |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          | |  |   |||     |  ||||| |||||  |||   |||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||| ||    || ||| |    ||            |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          |      |     |    |   ||  ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837
          ||  |   |||  |  | || || || || ||||| ||    ||    |||| |||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897
          |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957
            ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    958 AAG 960
            ||
Db   1095 GAG 1097


RESULT 5
US-08-525-058A-13
; Sequence 13, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-13

  Query Match             10.5%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 3.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
          || ||| ||||| |||||| ||| |  | || | |  ||  |  |||| || |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
           |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429
          | ||||  |     |||  ||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489
          ||   ||  | ||| ||     |  || || || ||  || | ||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||     ||| || ||||           | | | |      |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          | |  |   |||     |  ||||| |||||  |||   |||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          ||||| ||    || ||| |    ||            |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          |      |     |    |   ||  ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777
          |||||  |||| || || ||  | |||   || |||||||| || |||||  |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914
```

```
Qy    778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897
          |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957
            ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    958 AAG 960
            ||
Db   1095 GAG 1097
```

```
RESULT 6
US-08-696-731-13
; Sequence 13, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-13

  Query Match           10.5%;  Score 113.4;  DB 2;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 3.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;
```

```
Qy    196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
          ||  ||| |||||  ||||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
          ||      ||      |  ||| |  ||||  | |||||      ||||      |  | |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369
          || ||  |  | ||  |||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489
          ||   ||  | ||| ||    |  || || || ||  || | ||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
          |||     ||| || ||||          | | | |     |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
          |  |  |   |||      | |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
          |||||  ||    ||  |||  |    ||            |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
          |       |      |    ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777
          |||||  |||| || || ||  | |||   || ||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897
          |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957
            ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    958 AAG 960
            ||
Db   1095 GAG 1097
```

```
RESULT 7
US-09-042-531-13
; Sequence 13, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
```

```
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/042,531
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1654 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-13

  Query Match             10.5%;  Score 113.4;  DB 4;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 3.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     196 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 255
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     256 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 312
           |      ||      |  ||| |  |||| | |||||      ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     313 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 369
           || ||  |  | || ||||||||  |||  |   | |  |    |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     370 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 429
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     430 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 489
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     490 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 540
           |||     ||| || ||||           | | | |      |    | |       | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     541 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 600
           |  |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     601 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 660
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     661 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 720
           |      |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     721 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 777
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     778 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 837
           ||  |   |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     838 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 897
           |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     898 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 957
           |||  ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     958 AAG 960
           ||
Db    1095 GAG 1097


RESULT 8
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             10.4%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 6e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;
```

```
Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
          ||||| |||| | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
          | || |||| |   |   |  | ||    | || |       | ||||| |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
          ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
          |   ||  |   |  |     |          | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
          | |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
          |    ||           |  |  |  |    |||||  |   |              |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
          || |  |  |  ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          |||||  || |      ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 9
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             10.4%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 6e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
          ||||| |||| | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
          | || |||| |   |   |  | ||    | || |       | ||||| |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
          ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
          |   ||  |   |  |     |          | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
          | |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
          |    ||           |  |  |  |    |||||  |   |              |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
          || |  |  |  ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
```

```
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          |||||  || |       |||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 10
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             10.4%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 6e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
          | || ||||  |    |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
            ||||| ||| | || |||||  |||| || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
          |   ||  |    |  |    |     |          | |  |  |  |   |||

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
            |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
            |    ||          |  |  |  |     |||||  |   |  |           |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
          || |  |  |   ||||   ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
          ||  | |||   || ||||||||| || |||||  |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
          ||||| || |||||  || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          |||||  || |       |||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 11
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
```

```
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             10.4%;  Score 112.2;  DB 2;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 6e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
           |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db     271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy     331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
             | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db     331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy     388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db     391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy     448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy     508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
           |   ||  |   |  |   |   |          | |  |  |  |   |||
Db     511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy     559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
             |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db     571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy     619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
             |    ||           |  |  |  |    |||||  |   |                |
Db     631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy     679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
           || |  |  |  ||||   ||  | ||    |  || || || ||  |||| || ||
Db     691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy     739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
           ||  | |||   || ||||||| || |||||  |     || ||  |   |||  |  |
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy     796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
           || || || ||  |||||  ||    ||    |||||  ||||   |   | |||  | ||  |
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy     856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
           |||||  || |||||  || |   |    ||||  || || ||   ||| ||| ||||  |
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy     916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
           |||||  || |       |||||  || ||||||  |||| || ||
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 12
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match             10.4%;  Score 112.2;  DB 4;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 6e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
           |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db     271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy     331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
             | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db     331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy     388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db     391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy     448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy     508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
           |   ||  |   |  |   |   |          | |  |  |  |   |||
Db     511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy     559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
             |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db     571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy     619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
             |    ||           |  |  |  |    |||||  |   |                |
Db     631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690
```

```
Qy     679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
           || | |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db     691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy     739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
           ||  | |||   || |||||||| || ||||| |     || ||  |  |||  | |
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy     796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
           || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy     856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
           ||||| || ||||| || |   |    |||| || || ||   ||| ||| |||| |
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy     916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
           |||||  || |       ||||| || ||||||  |||| || ||
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 13
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             10.4%;  Score 112.2;  DB 3;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 8.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
```

```
           |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
           | || ||||  |   |  |  | ||    | ||  |      | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
           |||||  ||| | ||  |||||  ||||  ||  ||   |  ||   |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
           |  || || || ||  || | ||||||  | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
           |  ||  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
           |  |||||  |||||  |||   ||||||| || ||    |||||  ||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
           |    ||           |  |  |  |    ||||| |   |                 |
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
           || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
           ||  | |||   || |||||||| || ||||| |     || ||  |  |||  | |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
           || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
           ||||| || ||||| || |   |    |||| || || ||   ||| ||| |||| |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
           |||||  || |       ||||| || ||||||  |||| || ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             10.4%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 8.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
          ||||| |||| | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
          | || |||| |   |   | || |       | |||||      |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
            ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
          |   ||  |   | |   |     |           | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618
          |  |||||  |||||  |||  |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678
          |    ||           |  |  |  |    |||||  |   |                |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795
          ||  | |||   || ||||||||| || ||||| |      || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855
          || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915
          ||||| || ||||| || |    |    ||||  || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960
          |||||  || |    |||||  || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 15
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             10.4%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 8.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    271 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 330
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    331 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 387
          | || ||||  |   |   | || |       | |||||      |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    388 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 447
            ||||| ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    448 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 507
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    508 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 558
```

```
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     559 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 618

Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     619 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 678

Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     679 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 738

Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     739 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 795

Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     796 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 855

Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     856 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 915

Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     916 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 960

Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:21:37
Job time : 29.0361 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 1146.59 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-4_COPY_1454_2368
Perfect score: 915
Sequence: 1 accaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description
	1	915	100.0	3019	9	AB023021	AB023021 Homo sapi
	2	913.4	99.8	80247	9	AL512406	AL512406 Human DNA
	3	910.2	99.5	2501	9	HSA238701	AJ238701 Homo sapi
	4	792	86.6	1128	10	AF345993	AF345993 Rattus no
	5	791.8	86.5	2139	10	AB015426	AB015426 Mus muscu
	6	790.4	86.4	2156	10	AB049819	AB049819 Rattus no
	7	783.8	85.7	1705	10	AF230460	AF230460 Cricetulu
	8	666	72.8	1707	5	AB035906	AB035906 Gallus ga
	9	581.2	63.5	1080	5	AB035905	AB035905 Xenopus l
	10	580.2	63.4	71396	2	AC100674	AC100674 Mus muscu
	11	277.8	30.4	1567	5	AB023627	AB023627 Danio rer
	12	234.6	25.6	1832	5	AB023628	AB023628 Danio rer
	13	222.6	24.3	1092	5	AB035907	AB035907 Xenopus l
	14	220.2	24.1	1264	5	AF288369	AF288369 Gallus ga
	15	176.2	19.3	2657	5	GGU73678	U73678 Gallus gall
	16	145.2	15.9	1429	10	CGU78737	U78737 Cricetulus
	17	145.2	15.9	1462	10	AF090450	AF090450 Cricetulu
	18	143.6	15.7	1718	10	AF090449	AF090449 Cricetulu
	19	137.2	15.0	1044	5	AB035908	AB035908 Gallus ga
	20	132	14.4	2293	4	BTA134FTG	X87810 B.taurus al
	21	132	14.4	12607	4	BTA132772	AJ132772 Bos tauru
c	22	128	14.0	71396	2	AC100674	AC100674 Mus muscu
	23	127.6	13.9	1158	10	AB039126	AB039126 Mus muscu
	24	127	13.9	1487	3	AF016899	AF016899 Schistoso
	25	127	13.9	1814	6	AR028686	AR028686 Sequence
	26	127	13.9	3594	10	MMU45980	U45980 Mus musculu
c	27	127	13.9	117993	2	AC091284	AC091284 Mus muscu
	28	125.6	13.7	1613	4	BTA132776	AJ132776 Bos tauru
	29	125.6	13.7	1619	4	BTA132775	AJ132775 Bos tauru
	30	125.6	13.7	1644	4	BTA132774	AJ132774 Bos tauru
	31	125.6	13.7	1873	4	BTA132773	AJ132773 Bos tauru
	32	124.6	13.6	1119	9	AF345881	AF345881 Macaca mu
	33	122.8	13.4	1170	10	AB039127	AB039127 Mus muscu
	34	122.8	13.4	1170	10	AB039131	AB039131 Mus muscu
	35	122.8	13.4	1615	10	MMU33457	U33457 Mus musculu
	36	122	13.3	1113	9	AF345884	AF345884 Saimiri s
	37	121.2	13.2	1170	10	AB039124	AB039124 Mus muscu
	38	121.2	13.2	1170	10	AB039125	AB039125 Mus muscu
	39	121.2	13.2	1170	10	AB039128	AB039128 Mus muscu
	40	121.2	13.2	1170	10	AB039129	AB039129 Mus muscu
	41	121.2	13.2	1170	10	AB039130	AB039130 Mus muscu
	42	121.2	13.2	1170	10	AB039132	AB039132 Mus muscu
	43	119.8	13.1	1119	9	AF345883	AF345883 Macaca mu
	44	118.4	12.9	1170	10	AB039133	AB039133 Mus spici
	45	118.2	12.9	1080	9	AB051859	AB051859 Pongo pyg

ALIGNMENTS

RESULT 1
AB023021
LOCUS AB023021 3019 bp mRNA linear PRI 23-JUN-1999
DEFINITION Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX, complete cds.
ACCESSION AB023021
VERSION AB023021.1 GI:5139692
KEYWORDS alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S., Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and Narimatsu,H.
TITLE Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX
JOURNAL FEBS Lett. 453, 237-242 (1999)
REFERENCE 2 (bases 1 to 3019)
AUTHORS Kaneko,M., Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
ORIGIN

  Query Match           100.0%;  Score 915;  DB 9;  Length 3019;
  Best Local Similarity  100.0%;  Pred. No. 8.6e-231;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 519

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 579

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 639

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 699

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 759

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 819

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 879

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 939

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 999

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 1059

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 1119

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 1179

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1180 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1239

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1240 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1299

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1300 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1359

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    1360 TGGTTTTGGAATTAA 1374


RESULT 2
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-504J9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-381A2 is at 80148 in this sequence.
            The true right end of clone RP11-77B15 is at 100 in this sequence.
```

```
FEATURES             Location/Qualifiers
     source          1. .80247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-504J9"
                     /clone_lib="RPCI-11.2"
     repeat_region   184. .1422
                     /note="L2 repeat: matches 875. .2419 of consensus"
     repeat_region   1744. .1834
                     /note="MIR repeat: matches 166. .258 of consensus"
     repeat_region   2950. .2977
                     /note="14 copies 2 mer ca 100% conserved"
     repeat_region   3411. .3669
                     /note="LTR16C repeat: matches 93. .364 of consensus"
     repeat_region   4372. .4704
                     /note="MLT1A1 repeat: matches 5. .365 of consensus"
     misc_feature    complement(7821. .8406)
                     /note="match: GSS: Em:AQ550748"
     repeat_region   8126. .8184
                     /note="MER58B repeat: matches 272. .332 of consensus"
     repeat_region   8185. .8696
                     /note="MLT2D repeat: matches 1. .513 of consensus"
     repeat_region   8779. .8817
                     /note="MLT2D repeat: matches 504. .542 of consensus"
     repeat_region   8819. .8974
                     /note="MER58A repeat: matches 6. .162 of consensus"
     misc_feature    complement(9443. .9891)
                     /note="match: GSS: Em:B90633"
     misc_feature    complement(9452. .9883)
                     /note="match: GSS: Em:AQ119695"
     repeat_region   9782. .9894
                     /note="MIR repeat: matches 92. .205 of consensus"
     repeat_region   10089. .10212
                     /note="MIR repeat: matches 135. .262 of consensus"
     misc_feature    complement(10093. .10671)
                     /note="match: GSS: Em:AQ237061"
     repeat_region   10223. .10282
                     /note="MIR repeat: matches 203. .262 of consensus"
     misc_feature    complement(10315. .10631)
                     /note="match: GSS: Em:AQ200259"
     misc_feature    10805. .11186
                     /note="match: GSS: Em:AQ072353"
     repeat_region   11657. .11844
                     /note="L2 repeat: matches 2546. .2750 of consensus"
     misc_feature    complement(11659. .12084)
                     /note="match: GSS: Em:AQ518361"
     misc_feature    complement(11674. .11987)
                     /note="match: GSS: Em:AQ094678"
     misc_feature    12083. .12588
                     /note="match: GSS: Em:AQ691235"
     repeat_region   12596. .12810
                     /note="MIR repeat: matches 23. .248 of consensus"
     repeat_region   12920. .12986
                     /note="L1MA5 repeat: matches 6228. .6295 of consensus"
     repeat_region   12987. .13169
                     /note="L1MA5 repeat: matches 5852. .6043 of consensus"
     repeat_region   13168. .13280
                     /note="L1MA8 repeat: matches 5710. .5832 of consensus"
     repeat_region   13851. .17218
                     /note="L1PB1 repeat: matches 2787. .6155 of consensus"
     repeat_region   18903. .18953
                     /note="MLT2CB repeat: matches 439. .501 of consensus"
     repeat_region   18984. .19433
                     /note="MLT2CB repeat: matches 1. .460 of consensus"
     misc_feature    complement(19108. .19690)
                     /note="match: GSS: Em:AQ563935"
     repeat_region   20683. .20801
                     /note="AluJo/FRAM repeat: matches 169. .288 of consensus"
     repeat_region   20802. .20862
                     /note="AluJ/FLAM repeat: matches 3. .66 of consensus"
     repeat_region   20912. .21033
                     /note="MIR repeat: matches 114. .237 of consensus"
     repeat_region   21047. .21128
                     /note="41 copies 2 mer ta 63% conserved"
     repeat_region   21676. .21859
                     /note="L2 repeat: matches 2130. .2308 of consensus"
     repeat_region   21890. .23612
                     /note="L1P repeat: matches 841. .2570 of consensus"
     repeat_region   23601. .26609
                     /note="L1PA5 repeat: matches 3107. .6145 of consensus"
     repeat_region   26620. .26789
                     /note="MER5A repeat: matches 1. .174 of consensus"
     repeat_region   28521. .28627
                     /note="MER39 repeat: matches 5. .111 of consensus"
     repeat_region   28656. .28837
                     /note="MER39b repeat: matches 54. .235 of consensus"
     repeat_region   28838. .29142
                     /note="AluSc repeat: matches 4. .308 of consensus"
     repeat_region   29143. .29494
                     /note="MER39b repeat: matches 235. .549 of consensus"
     repeat_region   30529. .31344
                     /note="L1PA3 repeat: matches 5331. .6146 of consensus"
     repeat_region   32234. .32340
                     /note="L2 repeat: matches 2627. .2750 of consensus"
     repeat_region   32399. .32578
                     /note="MIR repeat: matches 12. .203 of consensus"
     repeat_region   32816. .32979
                     /note="MIR repeat: matches 91. .259 of consensus"
     misc_feature    complement(33006. .33450)
                     /note="match: GSS: Em:AQ261227"
     misc_feature    complement(33058. .33444)
                     /note="match: GSS: Em:AQ040453"
     repeat_region   33312. .33596
                     /note="L1M4 repeat: matches 3473. .3773 of consensus"
     repeat_region   34856. .34915
                     /note="30 copies 2 mer aa 68% conserved"
     misc_feature    complement(34987. .35547)
                     /note="match: GSS: Em:AQ321598
                     match: STS: Em:G55571"
     misc_feature    complement(36488. .36936)
                     /note="match: GSS: Em:B41073"
     misc_feature    complement(36654. .37090)
                     /note="match: GSS: Em:AQ772584"
     misc_feature    37102. .37591
                     /note="match: GSS: Em:AQ489567"
     misc_feature    complement(37581. .38221)
                     /note="match: GSS: Em:AQ418233"
     repeat_region   37720. .37772
                     /note="L2 repeat: matches 2657. .2710 of consensus"
     misc_feature    complement(37778. .38215)
                     /note="match: GSS: Em:AQ819773"
     misc_feature    38218. .38732
                     /note="match: GSS: Em:AQ599868"
     repeat_region   38861. .39164
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   39612. .39653
                     /note="21 copies 2 mer ac 90% conserved"
     repeat_region   40854. .40909
                     /note="L2 repeat: matches 2655. .2710 of consensus"
     repeat_region   42169. .42270
                     /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   42411. .42548
                     /note="MER5B repeat: matches 21. .181 of consensus"
     repeat_region   42780. .42938
                     /note="MER5B repeat: matches 17. .177 of consensus"
     repeat_region   43229. .43326
                     /note="MIR repeat: matches 40. .148 of consensus"
     repeat_region   45383. .45428
                     /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                     /note="match: STS: Em:G51995
                     match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                     /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
```

```
                /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match             99.8%;  Score 913.4;  DB 9;  Length 80247;
  Best Local Similarity   99.9%;  Pred. No. 2.2e-230;
  Matches  914;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35724 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 35783

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35784 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 35843

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35844 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 35903

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35904 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 35963

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35964 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 36023

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36024 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 36083

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36084 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 36143

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36144 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 36203

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36204 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 36263

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36264 TCTACTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 36323

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36324 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 36383

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36384 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 36443

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36444 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 36503

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36504 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 36563

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36564 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 36623

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db   36624 TGGTTTTGGAATTAA 36638


RESULT 3
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match             99.5%;  Score 910.2;  DB 9;  Length 2501;
  Best Local Similarity   99.7%;  Pred. No. 1.6e-229;
  Matches  912;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 329

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
```

```
Db    330 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 389

Qy    121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    390 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGCTGG 449

Qy    181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          |||||||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db    450 GATCTGACTAACTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 509

Qy    241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    510 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 569

Qy    301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    570 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 629

Qy    361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    630 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 689

Qy    421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    690 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 749

Qy    481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    750 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 809

Qy    541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    810 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 869

Qy    601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    870 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 929

Qy    661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 989

Qy    721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    990 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1049

Qy    781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1050 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1109

Qy    841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1169

Qy    901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db   1170 TGGTTTTGGAATTAA 1184


RESULT 4
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN

  Query Match           86.6%;  Score 792;  DB 10;  Length 1128;
  Best Local Similarity  91.8%;  Pred. No. 2.5e-198;
  Matches  837;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy      4 AAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACC 63
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    217 AAAACTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACC 276

Qy     64 TTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGAC 123
          ||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db    277 TTTGACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGAC 336

Qy    124 CGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGAT 183
          || ||  ||||||||||||||||||| || ||||| || ||| |||||||||| ||||||
Db    337 CGCTCGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGAT 396

Qy    184 CTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTG 243
          ||||| || |||||||||||| || ||||| ||||| ||||||||||||||||||||| ||
Db    397 CTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTA 456

Qy    244 GAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTG 303
          || ||||| |||||||| |||||||||||||||||||||||||||| ||||||||||
Db    457 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTA 516

Qy    304 ACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAAT 363
          ||||| |||||||||||||||| |||||||||| |||||||||||||| |||||||
Db    517 ACTTATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGC 576

Qy    364 CCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGG 423
          ||||| |||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db    577 CCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGG 636

Qy    424 AACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCAT 483
          || || |||||||||||| |||||||||||||| ||||| ||||| || ||||||||||
Db    637 AATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 696

Qy    484 ACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCT 543
          ||||| |||||||||||||||||||||||| |||||||||||| ||||||| |||||||||
```

```
Db    697 ACCTATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCT 756

Qy    544 GCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAG 603

Db    757 ACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 816

Qy    604 CTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAAC 663

Db    817 CTCTACAATGCATTTTTGGCTGGTTCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAAC 876

Qy    664 TATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGT 723

Db    877 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGT 936

Qy    724 GAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTT 783

Db    937 GAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTC 996

Qy    784 AACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTG 843

Db    997 AACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTG 1056

Qy    844 GCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGG 903

Db   1057 GCATGCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGG 1116

Qy    904 TTTTGGAATTAA 915

Db   1117 TTTTGGAATTAA 1128
```

```
RESULT 5
AB015426
LOCUS       AB015426               2139 bp    mRNA    linear   ROD 23-JUN-1999
DEFINITION  Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete
            cds.
ACCESSION   AB015426
VERSION     AB015426.1  GI:3702718
KEYWORDS    Fut9; alpha1,3-fucosyltransferase IX.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K.
            and Narimatsu,H.
  TITLE     Expression cloning and characterization of a novel murine alpha1,
            3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15)
            epitope in brain and kidney
  JOURNAL   J. Biol. Chem. 273 (41), 26729-26738 (1998)
  MEDLINE   98434588
REFERENCE   2  (bases 1 to 2139)
  AUTHORS   Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN
```

```
  Query Match           86.5%;  Score 791.8;  DB 10;  Length 2139;
  Best Local Similarity   91.6%;  Pred. No. 2.8e-198;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy      1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60

Db    280 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 339

Qy     61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120

Db    340 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 399

Qy    121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180

Db    400 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 459

Qy    181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240

Db    460 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 519

Qy    241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300

Db    520 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 579

Qy    301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360

Db    580 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 639

Qy    361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420

Db    640 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 699

Qy    421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480

Db    700 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 759

Qy    481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540

Db    760 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 819

Qy    541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600

Db    820 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 879

Qy    601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660

Db    880 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 939

Qy    661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720

Db    940 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 999

Qy    721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780

Db   1000 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 1059

Qy    781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840

Db   1060 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1119

Qy    841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900

Db   1120 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1179
```

```
Qy    901 TGGTTTTGGAATTAA 915
          ||||||||||||||||
Db   1180 TGGTTTTGGAATTAA 1194


RESULT 6
AB049819
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN

  Query Match           86.4%;  Score 790.4;  DB 10;  Length 2156;
  Best Local Similarity   91.7%;  Pred. No. 6.5e-198;
  Matches  836;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      4 AAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACC 63
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    514 AAAACTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACC 573

Qy     64 TTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGAC 123
          |||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db    574 TTTGACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGAC 633

Qy    124 CGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGAT 183
          || ||  |||||||||||||||| || ||||| || ||| |||||||||| ||||||
Db    634 CGCTCGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGAT 693

Qy    184 CTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTG 243
          ||||| || |||||||||| || ||||| ||||| ||||||||||||||||||| ||
Db    694 CTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTA 753

Qy    244 GAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTG 303
          || ||||| |||||||| |||||||||||||||||||||||||||||| ||||||||||
Db    754 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTA 813

Qy    304 ACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAAT 363
          ||||| |||||||||||||||| |||||||||| |||||||||||||| |||||||
Db    814 ACTTATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGC 873

Qy    364 CCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGG 423
          ||||| |||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db    874 CCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGG 933

Qy    424 AACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCAT 483
          || || |||||||||| |||||||||||||| ||||| ||||| || ||||||||||
Db    934 AATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 993

Qy    484 ACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCT 543
          ||||| |||||||||||||||||||||| |||||||||||| ||||||| |||||||||
Db    994 ACCTATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCT 1053

Qy    544 GCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAG 603
          |||||||||||||||||||||||||||||| ||||| |||||||||||| ||||||
Db   1054 ACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 1113

Qy    604 CTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAAC 663
          || |||||||| ||| ||||||| || |||||||||| ||||||||||||||||||||
Db   1114 CTCTACAATGCATTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAAC 1173

Qy    664 TATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGT 723
          |||||||||||||||||||| |||||||||||||||||||||||| |||||| ||||||
Db   1174 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGT 1233

Qy    724 GAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTT 783
          ||| | ||||| |||||||||| || ||||||||||||||||||| ||||||||||
Db   1234 GAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTC 1293

Qy    784 AACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTG 843
          |||||||||||||||| |||||||| || ||||| ||||||||||||||||||  ||
Db   1294 AACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTG 1353

Qy    844 GCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGG 903
          || |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db   1354 GCATGCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGG 1413

Qy    904 TTTTGGAATTAA 915
          ||||||||||||
Db   1414 TTTTGGAATTAA 1425


RESULT 7
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
```

```
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPPLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN

  Query Match             85.7%;  Score 783.8;  DB 10;  Length 1705;
  Best Local Similarity   91.0%;  Pred. No. 3.6e-196;
  Matches  833;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          || ||||||||||||||||||||||| |||||||| || ||||||||||||||||||||||
Db     227 ACAAAAACTGATTATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAG 286

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          ||||||||||||||||||||||||| ||||||||||||||| |||||||||||| ||
Db     287 ACCTTTGACCTTACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACA 346

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||| |||||||||||||||||||| || ||||| || ||| |||||||||| |||
Db     347 GACCGCTCACTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGG 406

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          |||||||| || |||||||||| |||||||| ||||| ||||||||||||||||||||
Db     407 GATCTGACTAACTTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAAC 466

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          | || || || ||||| || |||||||||||||||||||||||| ||||| |||||||||
Db     467 CTAGAGTCTCCCACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACT 526

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          || || || ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     527 CTAACATATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 586

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          |||||||| |||||||||||||||||||||||||  |||||||||||||||||||||||
Db     587 AATCCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAAC 646

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||| || |||||||||| |||||||||| |||||||| ||||| || |||||||||
Db     647 TGGAATCCGGAGCATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATC 706

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          |||||||| |||||||| |||||||||| || ||||| |||||| | ||||| ||||||
Db     707 CATACCTATGGGCAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATA 766

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||| |||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db     767 TCTACTTGTAAATTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAA 826
```

```
Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||| |||||||| ||| |||||||||| |||||||||||||||||||||| ||||||
Db     827 AAGCTCTACAATGCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAA 886

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     887 AACTATGAGAATTATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 946

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          |||||| | ||||| ||||| |||||||| |||||||||||||||| ||||||||||
Db     947 AGTGAGTTGGCAAAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAT 1006

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          || |||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db    1007 TTCAACTGGAGGAAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGC 1066

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          |||||||| ||||| || |||||||| ||||| |||||||||| ||||||||||||||||
Db    1067 CTGGCTTGTGATCACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAA 1126

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    1127 TGGTTTTGGAATTAA 1141


RESULT 8
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match             72.8%;  Score 666;  DB 5;  Length 1707;
```

```
  Best Local Similarity   83.0%;  Pred. No. 4.3e-165;
  Matches  759;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy       2 CCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGA 61
           |||||||||||| |||||||||||||| |||| |||| ||| | |||||||||||| ||||
Db     434 CCAAAACTGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGA 493

Qy      62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGG 121
           | || || || |||||||||||| | |||||||||||||| || |||||||| || |  |
Db     494 CATTCGATCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTG 553

Qy     122 ACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGG 181
           |||| ||||| || ||||| |||||||||||||| || |||||| | ||||| |  ||||
Db     554 ACCGCTCACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGG 613

Qy     182 ATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATT 241
           ||||||| |||||||||||||||| |||||||| |||||||||||||||||||||||| |
Db     614 ATCTGACTAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACT 673

Qy     242 TGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTC 301
           ||||||| |||||||| ||||| |||||||||||||||| ||| | |||||||| || |
Db     674 TGGAATCTCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCC 733

Qy     302 TGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAA 361
           |||||||||| |||||||||||||| || |||||||||||||| | |||  ||  | |||
Db     734 TGACTTACCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAG 793

Qy     362 ATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 421
              ||||    | |||||||||||| ||||| || ||||| || |||||||| |||||||
Db     794 GCGCCTTTACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACT 853

Qy     422 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 481
           ||||||||||||||||  |||||||||||||||||||||| |||||   ||||||||||
Db     854 GGAACCCTGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCC 913

Qy     482 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 541
           |||||||||||||||| |||||||| || ||||| |||||||| || ||||| || || |
Db     914 ATACCTACGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCT 973

Qy     542 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 601
           |   ||| ||||| || ||||| |||||||||||||||||||| |||||||| || || |
Db     974 CCAGTTGCAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGA 1033

Qy     602 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 661
           | || ||||||||| |||||||||| || ||||| ||||| ||||| || || || ||||
Db    1034 AACTTTACAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAA 1093

Qy     662 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 721
           |||| |||||||| |||||||||| || || || || |||||||||| |   || || |
Db    1094 ACTACGAGAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAA 1153

Qy     722 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 781
           | ||||| ||| | ||||   |    | |||||||| || ||| | ||||||||||||
Db    1154 GAGAGCTTGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACT 1213

Qy     782 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 841
           | |||||||||||||||||| | ||  |||||||   ||| |||||||| |||||||||
Db    1214 TCAACTGGAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTC 1273

Qy     842 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAAT 901
           | ||||| |||||||||||||| || || ||||| |||||  |||| |||||||| ||||
Db    1274 TTGCTTGTGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAAT 1333

Qy     902 GGTTTTGGAATTAA 915
           |||||||||| |||
Db    1334 GGTTTTGGAACTAA 1347


RESULT 9
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN

  Query Match           63.5%;  Score 581.2;  DB 5;  Length 1080;
  Best Local Similarity   77.2%;  Pred. No. 1e-142;
  Matches  706;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       2 CCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGA 61
           | |||    ||||||  ||||||||     ||||  | |||||||||||| ||||||||||
Db     167 CTAAACAGAATTATTACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGA 226

Qy      62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGG 121
           | ||||| || | |||||||||  || |||| || ||  | || |||||||| || || |
Db     227 CTTTTGAACTGAAATCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTG 286

Qy     122 ACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGG 181
           ||||  |||| |||||||| || ||||| || ||  | || ||| ||||||| || ||||
Db     287 ACCGCGCACTCTACAACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGG 346

Qy     182 ATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATT 241
           |||| |  ||  | || || ||    |||||||| |||||||||||||||||||||||
Db     347 ATCTCATTAACCTGCCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATC 406

Qy     242 TGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTC 301
           |||||||||||| || || || ||||| |||||||||||||||| ||||||||||||
Db     407 TGGAATCACCAACACATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTC 466

Qy     302 TGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAA 361
           |||||||| | ||||| |||||||||||||||||||||||||||| ||   ||||||||||
Db     467 TGACTTACAGGCGTGACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAA 526

Qy     362 ATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 421
           | || || |  ||||| |||||||||||||| || || || || |||||||| || ||||
Db     527 AACCATTTGACTTTGAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACT 586
```

```
Qy     422 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 481
           ||||||| || |||||||||||||| |||||||||| || | ||||  |||||||||
Db     587 GGAACCCAGAACATGCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTA 646

Qy     482 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 541
             || || |||||||| ||||| |||||  | |  |||||||   || |||| || ||
Db     647 TAACTTATGGGCAAGCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTG 706

Qy     542 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 601
           |  |||| |||||||||| || |||||||||| || ||||| ||||| ||||| ||||
Db     707 CATCTTGCAAATTTTATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAA 766

Qy     602 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 661
           | || || ||||| ||  |||||| || || ||  |||| |||||||  |||| ||||
Db     767 AACTTTATAATGCATTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAA 826

Qy     662 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 721
           |||||||||| || |||| |||||||| || || || |||||||||| |  |||||| |
Db     827 ACTATGAGAACTACATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAA 886

Qy     722 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 781
           | ||  |||| |  ||||         |  ||||| | |   |    ||||| || ||||
Db     887 GAGAATTAGCTGACCATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACT 946

Qy     782 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 841
           ||||||||||  |||| || || ||| | || |  || |||||||| |||||||| |
Db     947 ACAACTGGAGGAAACATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCT 1006

Qy     842 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAAT 901
           | || || || |||||||||||||||||||||| || || || || ||| ||||||| |
Db    1007 TAGCGTGTGACCATGTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGT 1066

Qy     902 GGTTTTGGAATTAA 915
           |||| ||||| |||
Db    1067 GGTTCTGGAACTAA 1080
```

```
RESULT 10
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1702: contig of 814 bp in length
            *     1703 1802: gap of      100 bp
            *     1803     2642: contig of 840 bp in length
            *     2643 2742: gap of      100 bp
            *     2743     3637: contig of 895 bp in length
            *     3638 3737: gap of      100 bp
            *     3738     4635: contig of 898 bp in length
            *     4636 4735: gap of      100 bp
            *     4736     5657: contig of 922 bp in length
            *     5658 5757: gap of      100 bp
            *     5758     6692: contig of 935 bp in length
            *     6693 6792: gap of      100 bp
            *     6793     7703: contig of 911 bp in length
            *     7704 7803: gap of      100 bp
            *     7804     8742: contig of 939 bp in length
            *     8743 8842: gap of      100 bp
            *     8843     9787: contig of 945 bp in length
            *     9788 9887: gap of      100 bp
            *     9888    10819: contig of 932 bp in length
            *    10820 10919: gap of      100 bp
            *    10920    11808: contig of 889 bp in length
            *    11809 11908: gap of      100 bp
            *    11909    12830: contig of 922 bp in length
            *    12831 12930: gap of      100 bp
            *    12931    13847: contig of 917 bp in length
            *    13848 13947: gap of      100 bp
            *    13948    14886: contig of 939 bp in length
            *    14887 14986: gap of      100 bp
            *    14987    15911: contig of 925 bp in length
            *    15912 16011: gap of      100 bp
            *    16012    16903: contig of 892 bp in length
            *    16904 17003: gap of      100 bp
            *    17004    17906: contig of 903 bp in length
            *    17907 18006: gap of      100 bp
            *    18007    18931: contig of 925 bp in length
            *    18932 19031: gap of      100 bp
            *    19032    19945: contig of 914 bp in length
            *    19946 20045: gap of      100 bp
            *    20046    20951: contig of 906 bp in length
            *    20952 21051: gap of      100 bp
            *    21052    21949: contig of 898 bp in length
            *    21950 22049: gap of      100 bp
            *    22050    22950: contig of 901 bp in length
```

```
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp
```

Query Match 63.4%; Score 580.2; DB 2; Length 71396;
Best Local Similarity 86.0%; Pred. No. 1.8e-142;
Matches 692; Conservative 0; Mismatches 100; Indels 13; Gaps 4;

```
Qy      1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          || |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||
Db  65550 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 65609

Qy     61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          |||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db  65610 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 65669

Qy    121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db  65670 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 65729

Qy    181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          |||||||| || |||||||||||| || |||||||||||| ||||||||||||||||||||
Db  65730 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 65789

Qy    241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db  65790 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 65849

Qy    301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  65850 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 65909

Qy    361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          |||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db  65910 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 65969

Qy    421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          |||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db  65970 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 66029

Qy    481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          || ||||| || |||||||| |||||||| || || |||||||||| ||||||| ||||||
Db  66030 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 66089

Qy    541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||| ||||||||||||||||||  |  ||| |      ||  |  | | |   | |
Db  66090 TCTACTTGTAAATTTTATCTTTCTCTGAAAACTCATTCACAAGATTACTCCCAAAAG--- 66146

Qy    601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
              | ||     ||| ||||||| || |||||||||||| ||||| ||||||||||||
Db  66147 ------TCCACATGCTTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 66200

Qy    661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTC--ATGTGGAAGATTATAACTCT- 717
          |||||||||||||||||||| ||||||||||||||    |||||||||| | |||||
Db  66201 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCCATCGTGGAAGATTTTCACTCTC 66260
```

```
Qy    718 CCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT-ATACCTTAG 776
          ||||||||| |||||||    ||||||||||||||||| ||||||||||  ||||||||
Db  66261 CCCAGTGAGTTAGCAAATATCCTGAAGGAAGTCGACAANAACAATAAGTTGGTACCTTAG 66320

Qy    777 TTACTTTAACTGGAGGAAGGATTTC 801
          ||||| ||||| |  ||| |  |||
Db  66321 TTACTCTAACTCGNAGAAAGGATTC 66345


RESULT 11
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN

  Query Match             30.4%;  Score 277.8;  DB 5;  Length 1567;
  Best Local Similarity   58.5%;  Pred. No. 1.2e-62;
  Matches  527;  Conservative    0;  Mismatches  362;  Indels   12;  Gaps    2;

Qy     18 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 77
          | | ||||| |   |  || | |||||||  || || ||       |||||| | |
Db    461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy     78 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 137
           ||     |  ||||||||||||| || || ||| | |||   ||| |     ||  | |
Db    521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy    138 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 197
          ||| | |||| |||| || | || ||| |||  | ||   ||          |||| ||
Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGG---------ATTTGCC 631

Qy    198 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 257
          || |   || | ||| | ||||| || ||||| |||  |||| |||| || || | | |
Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy    258 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 317
            | |  | ||| ||    ||  |  |  |||||| |||||    | || ||||   ||
Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy    318 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTT 374
          ||||||||| | |||||||||   | |  || |    |  | |    | |          ||
Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy    375 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 434
            | || |  || || ||| | || |||||| | || |||||||   |     | ||
Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy    435 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 494
          |     |||| || |||||| |    ||| |||| || |  |  |||| ||
Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy    495 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 554
            ||     |    || |||||| |  |||       |    ||||  |||||||||
Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy    555 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 614
            || |||||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy    615 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 674
          |||   |   ||  | || |||||||| || || || || || || |||||||||  | |
Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy    675 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 734
          ||| |||  ||| | ||||| |||||||| ||||       ||||||  || || ||  |
Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy    735 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 794
            | || ||      | ||| ||||| |  |     |||       || || |||||||| ||
Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy    795 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 854
          ||||  | |   ||    | ||  |||| ||      ||||||  |||  | |||| ||| |
Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy    855 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 914
          | | |||||| || |  | ||||   || ||     || |  |    ||||| |||  |||
Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy    915 A 915
          |
Db   1352 A 1352


RESULT 12
AB023628
LOCUS       AB023628               1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
```

```
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match           25.6%;  Score 234.6;  DB 5;  Length 1832;
  Best Local Similarity   55.5%;  Pred. No. 2.9e-51;
  Matches  500;  Conservative    0;  Mismatches  389;  Indels   12;  Gaps    2;

Qy      18 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 77
           | | || |  |   |  || | ||| ||   || ||  |     | ||||||||  | |
Db     619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy      78 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 137
             ||         ||  |||||||| || || ||  | |||   ||| |    |||  | |
Db     679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy     138 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 197
           ||| | |||| |||| || | || ||| |||| || ||          ||| |||| ||
Db     739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGC---------ACAGATTTTCC 789

Qy     198 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 257
           || |   |  | |||   |||| || ||||| |||  |||| |||| || || || ||
Db     790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy     258 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 317
             |||||      | |  |  |   ||||||  ||||    | ||||  | | ||
Db     850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy     318 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA---AATCCCTTCGTGTT 374
           ||||||||| | |||||||| ||      |   |    ||||    | |       |||
Db     910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy     375 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 434
           | || |  || || ||| | || |||||| | || |||||||   |     | ||
Db     970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy     435 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 494
             |     | |  | | | ||| |    |||  |||| |  |    ||||| ||
Db    1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy     495 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 554
             ||      |    | |||| | |  |||        |  | ||||   |||||||||
Db    1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy     555 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 614
           || || |||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db    1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy     615 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 674
           |||        ||  | || ||||| ||  | || ||  |  ||  | ||||||||  | |
Db    1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy     675 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 734
           || |||  |||  | ||||| |||||||| ||||      ||||||  || || ||  |
Db    1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy     735 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 794
           | || |       | ||| |||| |  |    |||       |  || ||||||||||
Db    1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy     795 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 854
             ||    |  |||   | ||  ||| ||     |||| | |||  |  ||| ||| |
Db    1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy     855 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 914
           | | |||||| || |  | ||||   | |||    || |  |    ||||| |||  |||
Db    1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy     915 A 915
           |
Db    1510 A 1510


RESULT 13
AB035907
LOCUS       AB035907               1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN

  Query Match           24.3%;  Score 222.6;  DB 5;  Length 1092;
  Best Local Similarity   57.3%;  Pred. No. 4.2e-48;
```

```
  Matches  484;  Conservative    0;  Mismatches  349;  Indels   12;  Gaps    4;

Qy      19 AATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 78
           ||  ||   ||| |  |  | |||     |||||||||  |           |||
Db     199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy      79 TGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAAC 138
           |||||     | |||||  ||   || |||||  |||||||  || |    || ||| |
Db     259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy     139 AAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCT 198
           |  |  | ||  || || |||| ||||| |||||   ||  ||| | |   |  | |||
Db     319 GAGGCAGACGCCATTATGTTCCACCACCGGGACAT---TTCTGATTTCAGTGACCTTCCT 375

Qy     199 CAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 258
           | |    | |    || ||| || ||||| ||||||| || ||||| ||  | ||
Db     376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy     259 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 318
           | |||      |    |  | | |  | | || |||  |||  || | ||| |    ||
Db     436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy     319 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 378
           || ||||||    |||| || ||||||||     |   ||  || |    |    | |
Db     496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy     379 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 438
           ||||  ||||    ||  ||||| ||||||  | || |||||||||    || || |
Db     553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy     439 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 498
           || ||| |||||||||| ||||| || ||    | | ||||||  |    ||||| |
Db     613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy     499 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 558
           | |||     || | || ||  | ||  |  |    ||| | || |  | ||||||||||
Db     670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy     559 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 615
           |  | ||||| |||||  | |||| ||||||||||||||||||||| |     || ||
Db     730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy     616 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 675
           |||    |   | |  | ||  | ||  |||| ||     |  | ||||| |||   |
Db     790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy     676 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 735
           || ||       || |||||||| |||| ||||    |   ||||  |||| || | |
Db     850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy     736 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
           || |||||| |   | || |||||||  | | || ||||  || || || |   |||| ||||
Db     910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy     796 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 855
           || |   ||  |  |  |    || |||||    |||   ||    || ||| |
Db     970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy     856 GTGAA 860
           || ||
Db    1030 GTCAA 1034


RESULT 14
AF288369
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN

  Query Match             24.1%;  Score 220.2;  DB 5;  Length 1264;
  Best Local Similarity   58.0%;  Pred. No. 1.8e-47;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    6;  Gaps    3;

Qy      28 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 87
           || || ||| |||||   ||||| || || ||   |||| |  | |    ||||  ||
Db     175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy      88 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
           | ||  ||| ||      |||||  |||| ||  ||||       || | | ||  |  ||
Db     235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 207
           || ||  | || |||||| | ||  |   ||||||       |   |  | | |    |
Db     295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy     208 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 267
           ||||  || | ||||   ||||| ||| | ||  |||| || |||| |||  |
Db     355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy     268 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 327
             ||| ||| || ||| | || |||||||||| || | |||||  | || || || |||
Db     415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy     328 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 387
               || |||||     ||  | |   ||   | |||       ||     | |||  |
Db     475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy     388 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 447
           ||     ||  |||||  |||||| || ||||| ||||  ||          | | || |||
Db     535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy     448 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 507
           || ||| | ||||| |  |||  |||      |  ||   |||||||  |||      |
Db     595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGC-GGCAACACTTGCC 653

Qy     508 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 567
             |||| |||   | |   |    |  ||| ||||    |  || || ||  |  |||||
Db     654 CCTGTCCATG--GACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy     568 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATAC---AATGCTTTTCTGGCT 624
           || || |||   ||||| || |||||||| || ||||| |     |||||  |     ||
```

```
Db    712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy    625 GGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCA 684
          ||| |  | || ||||||||||| ||| || | |||||||||||| | |   | ||  |
Db    772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy    685 GATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAG 744
          || || ||||| || || || || | |  |  | |  |||| || ||  |||| ||| ||
Db    832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy    745 GAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          ||| |   ||   |||   ||   ||||   | |||||  | ||| | | |
Db    892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942


RESULT 15
GGU73678
LOCUS       GGU73678               2657 bp    DNA     linear   VRT 20-DEC-1996
DEFINITION  Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete
            cds.
ACCESSION   U73678
VERSION     U73678.1  GI:1657998
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N.,
            Schultz,T.L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
  TITLE     Molecular cloning and characterization of CFT1, a developmentally
            regulated avian alpha(1,3)-fucosyltransferase gene
  JOURNAL   J. Biol. Chem. 271 (51), 32960-32967 (1996)
  MEDLINE   97115837
REFERENCE   2  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M. and Larsen,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical
            Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES             Location/Qualifiers
     source          1. .2657
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            170. .1240
                     /gene="CFT1"
     CDS             170. .1240
                     /gene="CFT1"
                     /note="similar to mammalian fucosyltransferase FucTIV"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase"
                     /protein_id="AAC60060.1"
                     /db_xref="GI:1657999"
                     /translation="MELGPRWSPAARPGCPRRWRRRWALLGALLGAALALYVCVRELR
                     RRGSAAGRPEGEVTVLLWWEPFGRPWRPADCRRRYNITGCLLSADRGRYGEARAVLFH
                     HRDLALHGRQGLPRGPPPRPPRQRWVWMNFESPSHSPGLRGLAGLFNWTMSYRRDSDV
                     FVPYGYLYEPPSPRPFVLPRKSRLVAWVISNWNEEHARVRYYRQLKEHLPIDVYGARG
                     MALLEGSVVKTVSAYKFYLAFENSQHTDYITEKLWKNAFAASAVPVVLGPRRANYERF
                     IPADSFIHVDDFPSPRLLATYLKFLDKNKPSYRRYFAWRNKYEVHVTSFWDEHYCKVC
                     EAVRTAGNQLKTVQNLAGWFES"
BASE COUNT      581 a    712 c    818 g    546 t
ORIGIN

  Query Match           19.3%;  Score 176.2;  DB 5;  Length 2657;
  Best Local Similarity   53.1%;  Pred. No. 7.3e-36;
  Matches  451;  Conservative    0;  Mismatches  383;  Indels   15;  Gaps    3;

Qy     28 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 87
          ||  | ||| |||||  |  ||| || || |   |||    ||   |   |||||
Db    341 ACGGTGCTGCTGTGGTGGGAGCCCTTCGGCCGCCCCTGGCGCCCGGCCGACTGCCGGAGG 400

Qy     88 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             | |||||||   || ||||  || |   | |||||    | ||||  | |  | |

Db    401 CGCTACAACATCACGGGCTGCCTCCTGAGCGCCGACCGCGGGCGGTACGGCGAGGCTCGG 460

Qy    148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAG-- 205
          || || ||  |||| ||||| ||| |       | |         |   |  |  | |
Db    461 GCCGTGCTCTTCCACCACCGCGACCTGGCGCTGCACGGCCGCCAGGGGCTGCCCCGCGGG 520

Qy    206 -------CTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 258
                  |  | || ||    |||  ||| | |||||||| || || || ||  | |||
Db    521 CCCCCGCCGCGACCCCCGCGGCAGCGGTGGGTGTGGATGAACTTCGAGTCGCCCTCGCAC 580

Qy    259 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 318
          | |||     | | ||| | |       | || |||  |||  || | ||||| || ||
Db    581 TCCCCCGGGCTGCGGGGCCTGGCCGGGCTCTTCAACTGGACCATGTCGTACCGACGGGAC 640

Qy    319 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 378
          || ||  ||  ||||| || || | | |     |       |    ||       || |
Db    641 TCGGACGTCTTCGTGCCCTACGGGTACCTCTACGAGCCGCCGTCGCCGCGGCCCTTCGTG 700

Qy    379 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 438
          ||||  |||||       |||||  ||||||  | || ||||||||||   ||||| |||
Db    701 CTGCCTCGCAAAAGCCGGCTGGTGGCCTGGGTCATCAGCAACTGGAACGAGGAGCACGCC 760

Qy    439 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 498
          | ||    || |||    |||| |   |  | |    ||| |    ||||||   ||
Db    761 CGCGTGCGCTACTACCGGCAGCTGAAGGAGCACCTGCCCATCGACGTGTACGGG---GCG 817

Qy    499 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 558
          ||      |   ||    |  || |    ||  | || || |  || || ||
Db    818 CGCGGGATGGCGCTGCTGGAGGGCAGCGTGGTGAAGACGGTGTCGGCCTACAAGTTCTAC 877

Qy    559 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 615
          ||  |||| || || ||    |||| ||| |||||||| || ||||| |   | || ||
Db    878 CTGGCCTTCGAGAACTCCCAGCACACGGACTACATCACCGAGAAGCTGTGGAAGAACGCC 937

Qy    616 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 675
          ||    ||  || | || |||||||| || || ||    ||||  ||||| |||   |
Db    938 TTCGCCGCCAGCGCGGTGCCTGTTGTCCTCGGCCCCCGCAGGGCCAACTACGAGCGCTTC 997

Qy    676 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 735
          || || || || || ||||| || ||||| || |     |    ||  |  ||| || |
Db    998 ATCCCCGCCGACTCCTTCATCCACGTGGACGACTTCCCCAGCCCACGGCTGCTGGCCACC 1057

Qy    736 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          || ||||||    |||| ||||||||        |||   || || ||   |||| ||||
Db   1058 TACCTGAAGTTCCTCGATAAAAACAAGCCCAGCTACAGGAGGTATTTCGCCTGGCGGAAC 1117

Qy    796 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 855
          | |     ||  |  |  |    || |||||    |||   ||   || ||||||
Db   1118 AAGTATGAAGTCCACGTCACGTCTTTCTGGGATGAGCATTACTGCAAGGTTTGCGAGGCC 1177

Qy    856 GTGAAAAGG 864
          ||||  | |
Db   1178 GTGAGGACG 1186


Search completed: October  6, 2002, 04:33:24
Job time : 1341.59 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 102.682 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-4_COPY_1454_2368
Perfect score: 915
Sequence: 1 accaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	915	100.0	2676	21	AAZ92647	Human alpha-1,3-fu
2	915	100.0	2822	21	AAZ92646	Human alpha-1,3-fu
3	791.8	86.5	2036	21	AAZ92672	Murine alpha-1,3-f
4	791.8	86.5	2170	21	AAZ92645	Murine alpha-1,3-f
5	127.4	13.9	1814	18	AAT59505	Murine myeloid-lin
6	127	13.9	3594	18	AAT85219	Mouse alpha-fucosy
7	114.4	12.5	1701	15	AAQ77732	Human alpha-1,3-fu
8	114.4	12.5	1701	21	AAF21136	Human low adenosin
9	114.4	12.5	1701	21	AAA35014	Human adenosine re
10	114.4	12.5	6944	21	AAF21137	Human low adenosin
11	114.4	12.5	6944	21	AAA35015	Human adenosine re
12	113.4	12.4	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
13	113.4	12.4	1654	18	AAT61680	Human alpha(1,3)-f
14	112.2	12.3	1086	15	AAQ56912	Lewis blood group
15	112.2	12.3	2042	18	AAT80111	Fucosyltransferase
16	112.2	12.3	2043	12	AAQ13330	Glycosyltransferas
17	112.2	12.3	2043	15	AAQ56906	DNA encoding a gly
18	112.2	12.3	2043	18	AAT76769	Human alpha 1,3/4
19	112.2	12.3	2043	18	AAT61675	Human alpha(1,3/1,
20	111.8	12.2	1126	21	AAF21135	Human low adenosin
21	111.8	12.2	1126	21	AAA35013	Human adenosine re
22	111.8	12.2	1316	15	AAQ56911	DNA encoding a gly
23	111.8	12.2	1316	18	AAT61679	Human alpha(1,3)-f
24	103.6	11.3	795	17	AAT13798	Alpha-(1,3/1/4)-fu
25	102.4	11.2	2175	11	AAQ06691	CDX, a MILA involv
26	100.8	11.0	1256	21	AAF21134	Human low adenosin
27	100.8	11.0	1256	21	AAA35012	Human adenosine re
28	100.8	11.0	1400	13	AAQ31436	Encodes a HeLa cel
29	100.8	11.0	1488	15	AAQ56910	DNA encoding a gly
30	100.8	11.0	2175	12	AAQ14382	Clone 7.2 encoding
31	100.8	11.0	2175	24	AAS17082	Human cDNA clone 7
32	100.8	11.0	2861	12	AAQ14383	Clone 1 encoding 1
33	100.8	11.0	2861	21	AAF21133	Human low adenosin
34	100.8	11.0	2861	21	AAA35011	Human adenosine re
35	100.8	11.0	2861	24	AAS17083	Human cDNA clone 1
36	100.8	11.0	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
37	100.8	11.0	3647	15	AAQ56909	DNA encoding a gly
38	100.8	11.0	3647	18	AAT61678	Human alpha(1,3)-f
39	97.6	10.7	2134	18	AAT59506	Human myeloid deri
40	66.6	7.3	1578	23	ABL11867	Drosophila melanog
c 41	66.6	7.3	4229	23	ABL11866	Drosophila melanog
42	52.2	5.7	1353	21	AAZ51685	C. elegans alpha-1
43	44.4	4.9	676	22	AAS41344	cDNA encoding nove
44	44.4	4.9	1479	24	ABA05334	Human fucosyltrans
45	44.4	4.9	2557	24	ABA05333	Human fucosyltrans

ALIGNMENTS

RESULT 1
AAZ92647
ID AAZ92647 standard; cDNA; 2676 BP.
XX
AC AAZ92647;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 295..1374
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

```
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match             100.0%;  Score 915;  DB 21;  Length 2676;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-255;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 519

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 579

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 639

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 699

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 759

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 819

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 879

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 939

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 999

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 1059

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 1119

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 1179

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1180 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1239

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1240 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1299

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1300 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1359

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db    1360 TGGTTTTGGAATTAA 1374


RESULT 2
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag=  a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
```

PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Claim 4; Page 134-142; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC AAZ92646 and AAZ92647 represent cDNAs encoding human
CC alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

Query Match 100.0%; Score 915; DB 21; Length 2822;
Best Local Similarity 100.0%; Pred. No. 7.9e-255;
Matches 915; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1454 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 1513

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1514 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 1573

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1574 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 1633

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1634 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 1693

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1694 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 1753

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1754 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 1813

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1814 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 1873

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1874 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 1933

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1934 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 1993

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1994 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 2053

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2054 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 2113

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2114 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 2173

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2174 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 2233

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2234 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 2293

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2294 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 2353

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    2354 TGGTTTTGGAATTAA 2368
```

RESULT 3
AAZ92672
ID AAZ92672 standard; cDNA; 2036 BP.
XX
AC AAZ92672;
XX
DT 05-JUN-2000 (first entry)
XX
DE Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; murine; mouse; ss.
XX
OS Mus sp.
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Example 8; Page 162-164; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The

```
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

  Query Match            86.5%;  Score 791.8;  DB 21;  Length 2036;
  Best Local Similarity  91.6%;  Pred. No. 3.6e-219;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           || |||||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db     174 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 233

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           ||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||
Db     234 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 293

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||| ||| |||||||||||||||||||| || ||||| || ||  |||||||||| |||
Db     294 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 353

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db     354 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 413

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db     414 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 473

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     474 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 533

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db     534 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 593

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db     594 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 653

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db     654 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 713

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||||| |||
Db     714 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 773

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db     774 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 833

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           |||||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||
Db     834 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 893

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db     894 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 953

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db     954 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1013

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1014 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1073

Qy     901 TGGTTTTGGAATTAA 915
           ||||||||||||||| 
Db    1074 TGGTTTTGGAATTAA 1088


RESULT 4
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag=  a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.
XX
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
```

```
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match             86.5%;  Score 791.8;  DB 21;  Length 2170;
  Best Local Similarity   91.6%;  Pred. No. 3.7e-219;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           || ||||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||
Db     280 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 339

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           |||||||||||||||||||||||||||||||||| |||||||| |||||||||||||!
Db     340 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 399

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db     400 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 459

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db     460 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 519

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db     520 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 579

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 639

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db     640 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 699

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db     700 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 759

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db     760 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 819

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db     820 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 879

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db     880 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 939

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Db     940 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 999

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db    1000 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 1059

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           |||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db    1060 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1119

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1120 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1179

Qy     901 TGGTTTTGGAATTAA 915
           ||||||||||||||| 
Db    1180 TGGTTTTGGAATTAA 1194


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag=  a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
```

```
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             13.9%;  Score 127.4;  DB 18;  Length 1814;
  Best Local Similarity   52.1%;  Pred. No. 1.3e-26;
  Matches  334;  Conservative    0;  Mismatches  301;  Indels    6;  Gaps    2;

Qy     104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
           | |||| ||| |   | ||||    |||     ||   | |||| ||  | ||||
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |          |   |     |||| | | |    |||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
           ||| | |||       |||||||| || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           | | || |||  |   |||| || || ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
           ||||  |     | |   ||   |   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           ||||||  | || || |   |     |||     | |  |||  |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
           |       |  |  |  ||   | ||| |  ||    |||        |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
           | || | || ||   |  |     |      || || ||  ||||||| || |||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           ||| |||||||| || ||| | |    ||||||  |   ||||||  |||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||  || |||  | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 741
           |||| || |  | |||| || |||| || ||    |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag=  a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag=  b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag=  c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag=  d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag=  e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag=  f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag=  g
FT                   /number= 3
FT   exon            2067..3552
FT                   /*tag=  f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             13.9%;  Score 127;  DB 18;  Length 3594;
  Best Local Similarity   52.3%;  Pred. No. 2.4e-26;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
           | |||| ||| |   | ||||    |||     ||   | |||| ||  | ||||
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336
```

```
Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |          |   |      |||| | | |    |||||||  |   |||
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
           |||  | |||        |||||||  || |  |  || |           | ||  |   |
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           |  |  ||  |||   |   ||||  ||  ||  ||||||||||||||||   ||  ||  ||  ||
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
           ||||  |     | |    ||   |    |    | | ||   ||||    |   |||
Db    2517 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           ||||||  | || ||  |   |      |||     | |  |||   |||    ||||
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
           |     |  |  |  ||   | ||| |  ||    |||       |    | |||
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
           | || | ||  ||  |  |    |     || ||  ||  |||||||  ||  |||   ||
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           |||  |||||||| ||  |||  | |    ||||||   |   ||||||  |||||||  || |
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||  ||  |||  | |||  |||  ||  | | |||  |||||  | ||  | ||  |
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           ||||  ||  |  | ||||  ||  ||||  ||  ||
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904
```

RESULT 7
AAQ77732
ID AAQ77732 standard; cDNA to mRNA; 1701 BP.
XX
AC AAQ77732;
XX
DT 22-JUN-1995 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA.
XX
KW Alpha-1,3-fucosyltransferase; sialylated Lewis blood group antigen;
KW Lewis x; Lewis a; glycosylation; glycoprotein; glycolipid; ds.
XX
OS Homo sapiens (THP-1 cells).
XX
FH Key Location/Qualifiers
FT CDS 123..1151
FT /*tag= a
FT /product= alpha-1,3-fucosyltransferase
XX
PN WO9423021-A.
XX
PD 13-OCT-1994.
XX
PF 28-MAR-1994; 94WO-JP00496.
XX
PR 29-MAR-1993; 93JP-0069016.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Hanai N, Kurata K, Nishi T, Sasaki K;
XX
DR WPI; 1994-333183/41.
DR P-PSDB; AAR63215.
XX
PT Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
PT of active sugar chains, e.g. sialylated Lewis X
XX
PS Claim 3; Page 99-102; 155pp; Japanese.
XX
CC The sequence coding for alpha-1,3-fucosyltransferase was isolated
CC from a human THP-1 cell cDNA library. The enzyme catalyses the
CC addition of fucose via a 1,3-linkage to N-acetylglucosamine
CC contained in glycoproteins and glycolipids, e.g. for sialylation of
CC Lewis blood group antigens.
XX
SQ Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

Query Match 12.5%; Score 114.4; DB 15; Length 1701;
Best Local Similarity 50.8%; Pred. No. 7.6e-23;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | |||||  ||  |    |  ||||    |||      ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| |||||  ||  |            | |  |   |  | ||  | |       ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |     |||   ||| | |||       |||| || || |  |||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |      | | || |||  |   ||||  |||||  || ||  ||  || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
           |   |||    ||||||  ||  ||  ||||   |       |||     | | ||  |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||     ||||          |    ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |    |  |       || | ||  ||| |  |     |      || || ||  ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           ||  ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           |||| || || || ||||| ||  |  |||  | |||||||    |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 8
AAF21136
ID AAF21136 standard; DNA; 1701 BP.
XX
AC AAF21136;
XX
DT 14-MAR-2001 (first entry)

XX
DE Human low adenosine antisense oligonucleotide related sequence #2703.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 1701 BP; 290 A; 527 C; 589 G; 295 T; 0 other;

Query Match 12.5%; Score 114.4; DB 21; Length 1701;
Best Local Similarity 50.8%; Pred. No. 7.6e-23;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |           | | |  |  | || | |      ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |     |||   ||| | |||        |||| || || |  ||||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |       | | || |||  |   ||||  ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           |||||  ||||||  | ||  |     |      ||||    |       |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
              |   |||    |||||| || || ||||   |       |||     | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||          |     |  ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
             |   | |       || | || ||| |  |     |      || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           |||| || || || ||||| ||  |  |||  | |||||||    |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 9
AAA35014
ID AAA35014 standard; DNA; 1701 BP.
XX
AC AAA35014;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2703.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.

```
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 957-958; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match            12.5%;  Score 114.4;  DB 21;  Length 1701;
  Best Local Similarity  50.8%;  Pred. No. 7.6e-23;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | |||||  || |   |  ||||    |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |       | | || |||  |   ||||  ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           |||||  ||||||  | ||  |      |       ||||    |        |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
              |   |||     |||||| || || ||||    |       |||      | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||           |     |  ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |    |  |      || | || ||| |  |    |     || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           |||| || || || |||||  ||  |  |||  | |||||||   |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | |||||  ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973


RESULT 10
AAF21137
ID   AAF21137 standard; DNA; 6944 BP.
XX
AC   AAF21137;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2704.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1034-1036; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
```

```
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

  Query Match           12.5%;  Score 114.4;  DB 21;  Length 6944;
  Best Local Similarity  50.8%;  Pred. No. 1.4e-22;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |      |||   ||| | |||      |||| || || |  ||||| | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
             |   |||    |||||| || || ||||   |      |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||          |     |  ||  ||  ||  |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |    |  |      || | || ||| |  |     |     || || || ||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           ||||  || || || ||||| ||  |  |||  | |||||||    |    | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216


RESULT 11
AAA35015
ID   AAA35015 standard; DNA; 6944 BP.
XX
AC   AAA35015;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2704.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 958-960; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

  Query Match           12.5%;  Score 114.4;  DB 21;  Length 6944;
  Best Local Similarity  50.8%;  Pred. No. 1.4e-22;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
```

```
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |       | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           ||||| ||||||  | ||  |      |       ||||    |        |||||  |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
             |   |||     |||||| || || ||||   |      |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||           |     |  ||  ||  || |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |    |  |       || | || ||| |  |     |      || || || |||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           ||||  || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216


RESULT 12
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             12.4%;  Score 113.4;  DB 15;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.5e-22;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |     ||     |  ||| |  ||||  | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || |||| |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||           | | | |      |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
             |       |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
```

```
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    793 AAG 795
          ||
Db   1095 GAG 1097


RESULT 13
AAT61680
ID   AAT61680 standard; DNA; 1654 BP.
XX
AC   AAT61680;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..1208
FT                   /*tag=  a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13643.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 6; Page 291-292; 329pp; English.
XX
CC   A DNA clone (AAT61680) codes for an enzyme (AAW13643) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC   precursors but not type I glycoconjugates.  It was identified in a
CC   human lambda phage genomic DNA library using a probe obtd. from
CC   Lewis enzyme cDNA (see also AAT61675).  The Fuc-TVI DNA can be used
CC   to construct animal cell lines with specific capabilities with
CC   respect to post-translational modification of the oligosaccharides
CC   of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC   use in oligosaccharide prodn., to genotype individuals at this
CC   fucosyltransferase locus, and in the prodn. of chimeric
CC   fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             12.4%;  Score 113.4;  DB 18;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.5e-22;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
          || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
           |    ||     |  ||| |  ||||  | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
          ||   ||  | ||| ||     |  || || || ||  || | ||||||  | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
          |||     ||| || ||||           | | | |       |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
          |  |  |  |||     |  ||||| ||||| |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
          ||||| ||     || ||| |    ||          |  |  |  |  |||||  |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
          |       |       |       |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
          ||  |   |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
          |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
            ||| ||| |||  | |||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    793 AAG 795
          ||
Db   1095 GAG 1097


RESULT 14
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
DE   Lewis blood group fucosyltransferase DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
```

```
XX
DR   WPI; 1994-048874/06.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding a Lewis blood group fucosyl
CC   transferase (Fuc-TIII).  The enzyme produced by the DNA may be
CC   non glycosylated.  This prevents premature loss of enzyme activity.
CC   It can also be used in in vitro reactions to modify cell surface
CC   oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-11.
XX
SQ   Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

  Query Match             12.3%;  Score 112.2;  DB 15;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.7e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 15
AAT80111
ID   AAT80111 standard; cDNA; 2042 BP.
XX
AC   AAT80111;
XX
DT   01-DEC-1997  (first entry)
XX
DE   Fucosyltransferase III coding sequence.
XX
KW   P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW   calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW   thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW   inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW   disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW   adult respiratory distress syndrome; tumour metastasis;
KW   fucosyltransferase III; acetylglucosaminyltransferase;
KW   autoimmune disorder; inflammatory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9706176-A2.
XX
PD   20-FEB-1997.
XX
PF   02-AUG-1996;   96WO-US12820.
XX
PR   17-MAY-1996;   96US-0649802.
PR   03-AUG-1995;   95US-0510920.
PR   15-MAY-1996;   96US-0017794.
XX
PA   (OKLA ) UNIV OKLAHOMA STATE.
XX
PI   Cummings RD,  Mcever RP,  Moore KL;
XX
DR   WPI; 1997-154206/14.
XX
PT   New O-glycan(s) derived from P-selectin glyco:protein ligand
PT   (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT   for treating inflammatory or autoimmune disorders or tumours
XX
PS   Disclosure; Page 80-81; 96pp; English.
XX
CC   This sequence represents the coding sequence for the human
CC   fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC   acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC   expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC   mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC   protein expressed on the surfaces of activated platelets and endothelium
CC   in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC   high affinity P-selectin ligand produced by leukocytes. Binding of
CC   P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC   treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC   highly extended extracellular domain, which is a feature of mucin-like
CC   proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC   the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC   O-glycans, and O-glycanated products of the invention. The products are
CC   used for inhibiting the binding of P-selectin and other selectins to
CC   PSGL-1. They can be used for the treatment of inflammation, injury
CC   resulting from ischaemia and reperfusion, bacterial sepsis and
CC   disseminated intravascular coagulation, adult respiratory distress
CC   syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC   other autoimmune or inflammatory disorders. The products can also be used
CC   in the production of antibodies and in diagnostic applications.
XX
SQ   Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

  Query Match             12.3%;  Score 112.2;  DB 18;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 3.6e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402
```

```
Qy     166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
           | || |||| |   |   |  | ||    | || |       | ||||| |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
           |||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
           |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
           |   ||  |    | |    |     |           | |  |  |  |   |||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
           |  |||||  ||||| |||   |||||||| || ||    ||||| ||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
           |    ||          |  |  |  |    ||||| |   |                |
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
           || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
           ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
           || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
           ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
           |||||  || |       ||||| || ||||||  |||| || ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 5, 2002, 23:13:52
Job time : 155.682 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 766.105 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-4_COPY_1454_2368
Perfect score: 915
Sequence: 1 accaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	573.8	62.7	697	9	BB397507	BB397507 BB397507
2	492.8	53.9	607	10	BI102760	BI102760 602888338
3	422	46.1	716	9	AV725071	AV725071 AV725071
4	324.8	35.5	649	9	AV338075	AV338075 AV338075
c 5	307.4	33.6	579	10	BM021081	BM021081 ie75f03.x
c 6	286	31.3	1101	12	CNS059ST	AL327638 Tetraodon
c 7	184.2	20.1	570	10	BM082484	BM082484 fu25e08.x
c 8	177	19.3	1100	12	CNS0579Z	AL324368 Tetraodon
c 9	170.2	18.6	1016	12	CNS0429P	AL271222 Tetraodon
10	161.4	17.6	844	12	CNS01V8V	AL168808 Tetraodon
c 11	157.4	17.2	887	12	CNS02Q0S	AL208693 Tetraodon
12	155.4	17.0	651	10	BJ036855	BJ036855 BJ036855
c 13	151.6	16.6	666	10	BJ050300	BJ050300 BJ050300
c 14	151.2	16.5	711	12	CNS034XF	AL228012 Tetraodon
c 15	151	16.5	501	12	B41073	B41073 HS-1052-B2-
16	146.6	16.0	640	10	BJ073227	BJ073227 BJ073227
17	145.2	15.9	648	9	BB595164	BB595164 BB595164
18	136.2	14.9	199	9	BB398103	BB398103 BB398103
19	133	14.5	625	9	BB649967	BB649967 BB649967
20	123.8	13.5	767	9	BB640911	BB640911 BB640911
c 21	123.4	13.5	674	10	BE605503	BE605503 fl10h04.x
22	120.6	13.2	971	12	CNS047IV	AL278032 Tetraodon
c 23	120.4	13.2	615	10	BI326811	BI326811 STR00078
c 24	117.8	12.9	1023	12	CNS02ANG	AL188773 Tetraodon
25	102.8	11.2	487	10	BE863396	BE863396 UI-M-BH0-
c 26	100.2	11.0	702	10	BJ043578	BJ043578 BJ043578
27	93.8	10.3	863	12	CNS02ROF	AL210840 Tetraodon
28	90.4	9.9	720	10	BF180380	BF180380 601804564
29	90	9.8	551	10	BJ035701	BJ035701 BJ035701
30	87.4	9.6	601	10	BJ036679	BJ036679 BJ036679
31	87.4	9.6	605	10	BJ028804	BJ028804 BJ028804
c 32	87	9.5	1056	12	CNS01VL7	AL169252 Tetraodon
33	86.8	9.5	304	9	AU081478	AU081478 AU081478
34	84.2	9.2	638	10	BJ035697	BJ035697 BJ035697
35	83.6	9.1	1057	12	CNS03ADC	AL235065 Tetraodon
c 36	83	9.1	566	10	BJ090565	BJ090565 BJ090565
37	80.2	8.8	628	9	BB651409	BB651409 BB651409
c 38	79.4	8.7	776	12	CNS03RIO	AL257289 Tetraodon
39	78	8.5	568	10	BM021317	BM021317 ie75f03.y
40	78	8.5	838	10	BG432556	BG432556 602500710
41	76.4	8.3	655	10	BG538740	BG538740 602566977
c 42	75	8.2	878	12	CNS02V1D	AL215194 Tetraodon
43	71.6	7.8	590	10	BM082755	BM082755 fu25e08.y
44	70.4	7.7	582	9	AW233458	AW233458 fj32h01.y
45	70	7.7	505	10	BJ029303	BJ029303 BJ029303

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN
```

Query Match 62.7%; Score 573.8; DB 9; Length 697;
Best Local Similarity 91.4%; Pred. No. 4.4e-148;
Matches 619; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

```
Qy     239 ATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGA 298
           | || || ||||| |||||||| ||| |||||||||||| ||| |||||||| |||||||
Db       4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy     299 CTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCA 358
           |||| ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||
Db      63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy     359 CAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTA 418
           |||||||||| |||||||||||||||||| ||||| |||||||||||||||||||||||
Db     123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy     419 ACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAA 478
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||
Db     183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy     479 TCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCA 538
           |||| ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||
Db     243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy     539 TATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGG 598
           ||||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |
Db     303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy     599 AAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGG 658
           ||||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db     363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy     659 AAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTC 718
           |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||
Db     423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy     719 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTT 778
           |||||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db     483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy     779 ACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCAT 838
           |||||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||
Db     543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy     839 GTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 898
           |  |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db     603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy     899 AATGGTTTTGGAATTAA 915
           |||||||||||||||||
Db     663 AATGGTTTTGGAATTAA 679
```

```
RESULT 2
BI102760
LOCUS       BI102760                607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: 1  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN
```

Query Match 53.9%; Score 492.8; DB 10; Length 607;
Best Local Similarity 90.1%; Pred. No. 1.2e-125;
Matches 549; Conservative 0; Mismatches 58; Indels 2; Gaps 2;

```
Qy      58 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACA 117
           |||||||||||| |||||||||||||||||||||||||| |||||||| ||||||||||||
Db       1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy     118 ACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGT 177
           || ||||| ||| |||||||||||||||||| || ||||| || ||  ||||||||||
Db      61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy     178 TGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATG 237
           |||||||||| || |||||||||| || |||||||||| |||||||| |||||||
Db     120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy     238 AATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTG 297
           ||||| || ||||| |||||||| |||||||||||||||||||| |||||||| ||||||
Db     179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy     298 ACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGC 357
           ||||| ||||| |||||||||||||||||||||||||||||||||||||||||||| |||
Db     239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy     358 ACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGT 417
           |||||||||| |||||||||||||||||| ||||| ||||||||||||||||||||||
Db     299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy     418 AACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAA 477
           |||||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||
Db     359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy     478 ATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACC 537
           ||||| ||||| || |||||||| |||||||| || || ||||||||| ||||||| |||
Db     419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy     538 ATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACG 597
           |||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||
Db     479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy     598 GAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGG 657
           |||||||| |||||||| ||| ||||||| || |||||| ||| ||||| |||||||||
Db     539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy     658 GAAAACTAT 666
           |||||||||
Db     599 GAAAACTAT 607
```

```
RESULT 3
AV725071
LOCUS       AV725071                716 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION   AV725071
VERSION     AV725071.1  GI:10830091
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .716
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCBLB02"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      219 a    113 c    133 g    249 t      2 others
ORIGIN
```

```
  Query Match             46.1%;  Score 422;  DB 9;  Length 716;
  Best Local Similarity   100.0%;  Pred. No. 5e-106;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     494 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy     554 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy     614 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy     674 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy     734 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy     794 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy     854 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy     914 AA 915
           ||
Db     421 AA 422
```

```
RESULT 4
AV338075
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
```

```
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430407H24"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory bulb"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      209 a    105 c    117 g    218 t
ORIGIN

  Query Match             35.5%;  Score 324.8;  DB 9;  Length 649;
  Best Local Similarity   91.5%;  Pred. No. 4.1e-79;
  Matches  344;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy     540 ATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGA 599
           |||  |||||||  |||  ||||||  ||||||||  |||||  ||||||||||||||||  ||
Db       2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61

Qy     600 AAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGA 659
           ||||||  ||||||||  |||  |||||||  ||  ||||||||||  |||||  ||||||||||
Db      62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy     660 AAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCC 719
           ||||||||||||||||||||||||  ||||||||||||||||||||||||||  |||||||||
Db     122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy     720 CAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTA 779
           |||||||  |||||||  ||||||||||||||  ||||||||||||||||  ||||||||||
Db     182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy     780 CTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATG 839
           ||||||||||||  ||||||||  |||||||  ||  |||||  ||||||||||||||||||||
Db     242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy     840 TTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 899
             ||||  |||||  |||||  ||||||||||||||||||||||||||||||||||||||||
Db     302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy     900 ATGGTTTTGGAATTAA 915
           ||||||||||||||||
Db     362 ATGGTTTTGGAATTAA 377


RESULT 5
BM021081/c
LOCUS       BM021081                579 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021081
VERSION     BM021081.1  GI:16535437
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
```

/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 203 a 103 c 93 g 180 t
ORIGIN

Query Match 33.6%; Score 307.4; DB 10; Length 579;
Best Local Similarity 99.7%; Pred. No. 2.6e-74;
Matches 308; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    607 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy    667 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy    727 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy    787 TGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCT 846
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy    847 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy    907 TGGAATTAA 915
          |||||||||
Db    279 TGGAATTAA 271
```

RESULT 6
CNS059ST/c
LOCUS CNS059ST 1101 bp DNA linear GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 045P22 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL327638
VERSION AL327638.1 GI:8221227
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="045P22"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA045DH11A1-end : T3"
BASE COUNT 287 a 249 c 269 g 278 t 18 others
ORIGIN

Query Match 31.3%; Score 286; DB 12; Length 1101;
Best Local Similarity 59.1%; Pred. No. 2.7e-68;
Matches 537; Conservative 0; Mismatches 366; Indels 5; Gaps 3;

```
Qy     12 TTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCT 71
          | ||   ||    || | ||| ||| ||||| | ||||| || || || ||||| |||||
Db    941 TGATAAAAACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCTTCGACCT 882

Qy     72 TACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACT 131
          |     ||| | ||    |||  | |  | || ||  |  | ||  ||||| |    ||
Db    881 GAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACAGAAACCT 822

Qy    132 GTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAA 191
          |  ||||| ||  |||  ||| |  |||| ||| ||||||||    | || ||| | ||
Db    821 TTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACCTGTCCAA 762

Qy    192 TTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACC 251
          | ||       |   | ||  || |||||||||||| ||| |  |  |||| || ||
Db    761 CATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGGAGTCTCC 702

Qy    252 AACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCG 311
          | ||| |  |     |  ||| || |||  | |||| |||||||| |||| || ||
Db    701 CTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGAACTATCG 642

Qy    312 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGT 371
          |  ||||| || ||       ||  |  || |     ||  ||| || | |     |
Db    641 TCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATGCTGAGGA 582

Qy    372 G---TTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCC 428
          |   ||||     || |||||| | ||| |||| || ||| |||| || |||||||||||
Db    581 GGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACTGGAACCA 522

Qy    429 TGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTA 488
          || |||||  |||| || |  |||||  ||  |   |||     ||||  |||| || ||
Db    521 GGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCCACACGTA 462

Qy    489 CGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTG 548
          || | |||||| ||  || |  ||   ||| || |     | || |||||  |    ||
Db    461 TGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCGCCAGCTG 402

Qy    549 TAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATA 608
          || || || || || ||||| || || |||||||| || ||| | || ||||| || ||
Db    401 CAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAAAACTGTA 342

Qy    609 CAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGA 668
          |||  |  ||   |  ||  | || || || |||||||| ||| || |||| ||||| ||
Db    341 CAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGAACTACGA 282

Qy    669 GAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCT 728
          |||    || |  | |||| | || || || |||||||| | |  |||  |||  |||||
```

```
Db    281 GAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCAAGGAGCT 222

Qy    729 AGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 788

Db    221 TGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATTTTCANGT 162

Qy    789 GAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTG 848

Db    161 GGCGAAGACACTTTAAAGCAAGGGCAGTGGCATTTGCCCGAGGC-CACCTGTCAGGCTTG 103

Qy    849 CGATCATGTGAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAATGGTTTT 907

Db    102 TGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAGTGGTACT 43

Qy    908 GGAATTAA 915

Db     42 GGGGTTGA 35


RESULT 7
BM082484/c
LOCUS       BM082484                570 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223
            3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;,
            mRNA sequence.
ACCESSION   BM082484
VERSION     BM082484.1  GI:16929414
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN
```

```
  Query Match           20.1%;  Score 184.2;  DB 10;  Length 570;
  Best Local Similarity  59.2%;  Pred. No. 3.6e-40;
  Matches  337;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    1;

Qy    217 TTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGC 276

Db    570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy    277 ATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCT 336

Db    510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy    337 TATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAA 396

Db    450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy    397 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGAGTCAAG 447

Db    390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331

Qy    448 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 507

Db    330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271

Qy    508 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 567

Db    270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211

Qy    568 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 627

Db    210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151

Qy    628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687

Db    150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91

Qy    688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 747

Db     90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31

Qy    748 GTCGACAAAAACAATAAGTTATACCTTAG 776

Db     30 CTAAATAAAAATGACATTGCATATCACAG 2


RESULT 8
CNS0579Z/c
LOCUS       CNS0579Z                1100 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            007H05 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL324368
VERSION     AL324368.1  GI:9557250
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
```

```
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007H05"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA007CD03C1~end : T7"
BASE COUNT      314 a    228 c    252 g    304 t      2 others
ORIGIN

  Query Match             19.3%;  Score 177;  DB 12;  Length 1100;
  Best Local Similarity   55.4%;  Pred. No. 4.3e-38;
  Matches  428;  Conservative    0;  Mismatches  332;  Indels   13;  Gaps    4;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| | ||| | ||||| | ||| |||| |||| |  |  ||  |||
Db     773 ATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTTGCATACGACTCTTGTAAAGTTTAC 714

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | ||| ||  | || ||  | ||||||   || || || || ||| ||||| | || ||
Db     713 TACAATATTGACGGCTGTGAGCTCACAGATGATCGATCCCTTTACGACAAAGCGCAGGCT 654

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           |||||| ||   |||  |||||    ||| |||||  ||| || ||   | |  |  ||
Db     653 GTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGGAAACTTGCCAGTGGAGCCGCGG 594

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           |||   || ||||  ||||||||| |  |  ||||||| || |   ||||   |   |
Db     593 CCATATTTTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGATA 534

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           |||  | ||     | || ||| |||||||||  ||| |    ||||| || ||   |
Db     533 CCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGTA 474

Qy     331 GTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 387
           |    ||   |     | | |  |     | |    |       | |   ||||
Db     473 GCCAGGTACCCTCTTACCATCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGAG 414

Qy     388 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC------ATGCCAGA 441
           ||| | ||| | |  || ||| | || || ||    |   |            | | |||
Db     413 AAAAACAAACTTGCCTGTTGGATCGTTAGCAATGCTACATCCACTGGGACAGGTACAAGA 354

Qy     442 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 501
           |   | | ||| ||||| |||  |||| ||| | ||   |   |  ||    | |
Db     353 GCGCAATTTTATAATGAACTATCCAAACATATTAACATTAGTGTTTTTGGTGTGGTGTAC 294

Qy     502 GGAGAATATGTCAATGATAAAAATTTGATT---CCTACCATATCTGCTTGTAAATTTTAT 558
             ||          || |   ||  |  |   || ||| |  ||   |||||  |||||
Db     293 ACAGGGGTCAGGTTGGAGATTGATCAGTATTACCCGACCGTTGCTAGCTGTAAGTTTTAC 234

Qy     559 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTT 618
           | || ||||| || ||| | ||||| ||||||||||| || ||| |  ||     |  ||
Db     233 TTATCATTTGAGAACTCATTACACAAAGATTACATCACAGAGAAGGTCAACGNACCCCTT 174

Qy     619 CTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATT 678
             || ||  | || ||||| |||||||||||  | || |  |||||  || |  |    |
Db     173 GCAGCAGGAACCGTTCCTGTAGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTTC 114

Qy     679 CCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTAT 738
           ||  |||| || || ||||| || || |||| |     ||| |  |  || ||| || ||
Db     113 CCTTCAGACTCCTTTATTCACGTAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCAT 54

Qy     739 CTGAAGGAAGTCGACAA-AAACAATAAGTTATACCTTAGTTACTTTAACTGGA 790
           |||       | |||||   |  |  ||  |||| |  | |||||| | ||||
Db      53 CTGCTCTCCTTGGACAAGGNATGAGGAGGCATACATGCGCTACTTTGAGTGGA 1


RESULT 9
CNS0429P/c
LOCUS       CNS0429P                1016 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            075P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271222
VERSION     AL271222.1  GI:7993186
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1016)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1016
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="075P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG075DH01SP1~end :
                     PUC-Ori"
BASE COUNT      246 a    216 c    253 g    298 t      3 others
ORIGIN

  Query Match             18.6%;  Score 170.2;  DB 12;  Length 1016;
  Best Local Similarity   57.0%;  Pred. No. 3.2e-36;
  Matches  394;  Conservative    0;  Mismatches  283;  Indels   14;  Gaps    4;

Qy      32 TTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGT 91
           ||||| ||||| | ||||  | ||   |   |||||  | |     |||    |    |
Db     715 TTCTGCTGTGGTTCTGGCCCATCGGCACGGTGTTTGATTTCAAGGATTGCTCCACCCACT 656

Qy      92 TCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAG 151
           |||| ||  |  | || |||||  ||     ||| | || ||||||| |||| |  | | ||
Db     655 TCAAGATTGACAGCTG-CATCTGACGGACGACAGATCTCTGTACAGCAAAGCAAAGGGAG 597

Qy     152 TTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGC 211
           |  |  || | |||  |||||  |  ||| |      |  || ||   |   | | |
Db     596 TCATTTTCTACCACAAAGACACAACATGGCACACAGACAGCTTGCCCACGGGTCCACGCC 537

Qy     212 CACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGA 271
           | |  |||||||  ||| ||||| | ||| |||||||| |||||  ||||   |   ||
Db     536 CCCATTTCCAGAGGTGGGTTTGGTTTAATGTGGAATCCCCAACAAACACATACAGGAAAC 477
```

```
Qy    272 GTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAG 331
Db    476 CAGGTCTGGAAAACCTGTTCAACCTGACGCTGAGCTACAGGAGAGACGGCGGCATTGTGG 417

Qy    332 TGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTG---TTTGAAGTGCCAAGCA 388
Db    416 CGAGGAACGAGATGTCTATCCGAGACCCAGGTGACACCGAGGACTTTGTTCTGCCCAAGA 357

Qy    389 AAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATG---------CCA 439
Db    356 AAGACAAGACGGTGTGTTGGATCGTCAGCAACAATGACCCGAGAACAGGAACAGGCACCA 297

Qy    440 GAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCAT 499
Db    296 GAGACCAATACTACCGTGACCTGTCCAAACACATCAAGGTGGACGTTTGGGGCAAAGCCT 237

Qy    500 TTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATC 559
Db    236 TTGGGAATTTCCTGAGTTTTGAGGAATACTATCCAACAATTGCCAGCTGTAAATTCTATC 177

Qy    560 TTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTC 619
Db    176 TGTCTTTCGAGAACTCCATCCACAAAGACTACATGACGGAGAAGATCAACGGGCCCTTTG 117

Qy    620 TGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATAT-T 678
Db    116 TGGCTGGGTCAGTTCCTATAGTTTTGGGACCTCCAAAAAAAAAAAAAAAAAAAAATTCTAT 57

Qy    679 CCAGCAGATTCATTCATTCATGTGGAAGATT 709
Db     56 CCTTCCGATTCGTTTATTCACGTTAACGACT 26
```

RESULT 10
CNS01V8V
LOCUS CNS01V8V 844 bp DNA linear GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 198K09 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL168808
VERSION AL168808.1 GI:7806865
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 844)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 844)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 844)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .844
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="198K09"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG198AF05SP1~end : PUC-Ori"
BASE COUNT 199 a 190 c 205 g 245 t 5 others
ORIGIN

Query Match 17.6%; Score 161.4; DB 12; Length 844;
Best Local Similarity 56.2%; Pred. No. 8.3e-34;
Matches 319; Conservative 2; Mismatches 244; Indels 3; Gaps 1;

```
Qy     25 ACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAA 84
Db    280 ACCACGGTGCTGCTCTGGTTCTGGCCATTTGGGATTACCTTTGACCCGAACGTGTGCAGC 339

Qy     85 GCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCC 144
Db    340 GCTTACTATGGTATAGACGGCTGTTTCATTACAGCGGACAGGAACTATTTCAGCGAAGGA 399

Qy    145 CATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAA 204
Db    400 ATGCGGNTCGTTATTCATCATCGAGAMATTTCTCCCGACCTGTCCAATCTGCCGCAGCAG 459

Qy    205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
Db    460 GAGCGGCCGTTGTACCARATGTGGGTCTGGTTCAACATGGAGAATCCTTCGTACTCCGAC 519

Qy    265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
Db    520 AGGTTTCTTGGAATGGACGGCTTGTTCAACCTGACTCTGAATTATCGTCTGGATGCTGAT 579

Qy    325 ATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCA 384
Db    580 ATTTCCATGCCTCATGTGTACCTGGTTCCTGAAAAAAGTGAGGAAGACTTTGTCCCCCCC 639

Qy    385 AGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTC 444
Db    640 AGCAAAGATAAGTTGGTCTGCTGGATAGTCAGCAACTGGAAGGAGCATTATATACGGGTG 699

Qy    445 AAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGA 504
Db    700 AAATACTTCAAGGAGCTGTCCAAGCACATAAAGATCACCACTTATGGACGTGGCTTTGGG 759

Qy    505 GAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCC 564
Db    760 GGCTA---CAACTGGGGAAGTGCCTCGTCCATCATGGCCAGCTGTAAGTTCTACCTGTCT 816

Qy    565 TTTGAAAATTCAATCCACAAGGATTACA 592
Db    817 TTTGAGAACTCCATCTACATTGACTACA 844
```

RESULT 11
CNS02Q0S/c
LOCUS CNS02Q0S 887 bp DNA linear GSS 14-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 156L20 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL208693
VERSION AL208693.1 GI:7867512
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 887)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the

```
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 887)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .887
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="156L20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG156DF10LP1~end : T7"
BASE COUNT      232 a    188 c    206 g    253 t      8 others
ORIGIN

  Query Match             17.2%;  Score 157.4;  DB 12;  Length 887;
  Best Local Similarity   54.2%;  Pred. No. 1.1e-32;
  Matches  380;  Conservative    5;  Mismatches  301;  Indels   15;  Gaps    3;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| | ||| | ||||| | ||| |||| || |         || ||  |||
Db     703 ATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTGGCATACGACTCTTGTAAAGTTGAC 644

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           : ||| ||  |  | ||  | ||:|||   || || || ||  || ||||| | || ||
Db     643 KACAATATTGACAGCTGTGAGCTSACAGATGATCGATCCCTTGACGACAAAGCGCAGGCT 584

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           |||||| ||   |||   |||||    ||| ||||| |||| || ||   | |  |  ||
Db     583 GTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGCAAACTTGCCAGGGGAGCCGCGG 524

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           |||   |: ||||  |||||||||| |  |  ||||||| || |   ||||   |   |
Db     523 CCATATTKTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGATA 464

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
              ||| | ||      | || ||| |||||||||  ||| |    ||||| || ||
Db     463 CCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGTG 404

Qy     331 GTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 387
           |    ||   |    | | |  |     | |    |       | |   ||||
Db     403 GCCAGGTACCCTCTTACCATCAGAGAGGAAAGACTGACGGAGAAAATAGKTCTGCCGGAG 344

Qy     388 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCA-- 445
           ||| | ||| | |||||||||| | || || ||           |:  || ||  | ||
Db     343 AAAAACAAAATTGTGTGCTGGATCGTCAGCAAAACTGCAAAAAASACTGGCACGGGCACA 284

Qy     446 -------AGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 498
                 | |  |    ||| ||| |||    ||| |  |  ||   |  ||    ||
Db     283 AGGGAACAATTCTTTGGTGAACTATCCAAGCATATTCATGTTGATGTTTTTGGTGTGGCG 224

Qy     499 TTT---GGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
              ||||   |  |  |  | |  |     || ||||  ||   ||||| |||
Db     223 GACACAGGAGTGAAGTTGGAGATTGATCAGTATTACCCGACCATTGCTAGCTGTAAGTTT 164

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 615
           || || |||||||| || |||   ||||| |||:|||||| || ||| |  ||    |
Db     163 TACCTGTCCTTTGAGAACTCATAACACAAAGATKACATCACAGAGAAGGTCAACGGACCC 104

Qy     616 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 675
            ||   |  ||  | || ||||| ||||||||||  | || |  ||||| || |  |
```

```
Db     103 CTTGCAGTAGGAACCGTTCCTGTGGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTC 44

Qy     676 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTC 716
             | ||  |||| || || ||||| || || |||| |   ||
Db      43 TTCCCTTCAGACTCCTTAATTCACGTAGATGATTTTCTTTC 3


RESULT 12
BJ036855
LOCUS       BJ036855                651 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  BJ036855 NIBB Mochii normalized Xenopus neurula library Xenopus
            laevis cDNA clone XL037k06 5', mRNA sequence.
ACCESSION   BJ036855
VERSION     BJ036855.1  GI:17398869
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL037k06"
                     /clone_lib="NIBB Mochii normalized Xenopus neurula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 15"
BASE COUNT      191 a    157 c    119 g    181 t      3 others
ORIGIN

  Query Match             17.0%;  Score 155.4;  DB 10;  Length 651;
  Best Local Similarity   56.6%;  Pred. No. 3.5e-32;
  Matches  329;  Conservative    0;  Mismatches  246;  Indels    6;  Gaps    2;

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || ||    ||  | ||| ||||| | |  ||||| || ||||  |||| |||| | |
Db      74 CCCCCAAAACAATATTGGGTTTGGTTTACCTTGGAGTCCCCAAGCCACAATCCCAACACA 133

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
              || ||| |  |  | || |||||  || |||||||   ||| ||||||||
Db     134 GCCTTTATGGAGAAACTCATCAATCTGACCATGTCTTACCGTGCTGACTCAGATATTTTT 193

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           ||||||||||  |                    ||      ||||   |||
Db     194 ACTCCTTATGGCTGGCTAGAAAGGCATGATGGAAAAGAAAACTTCACTATGCCCGCCAAG 253

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
           ||| | ||   |||||| || ||||||||||||||  |       ||| |||| |||
Db     254 ACAAAACTAGTAGCCTGGGTAGTCAGTAACTGGAACCCAAACTCCAGAAGAATCAAATAT 313

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           ||  | ||| |||     |  | |||   |  ||| ||| || | |     | |     |
Db     314 TATGAGGAGTTAAAGCCACACCTTGCTGTTGATATCTATGGCCTACAGCATTTACCGCTT 373

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |     ||| |    |||   |  | ||    |  |||||||| |  | ||||||
Db     374 CCTAGACCAAAACA---CATTGAAATTCTTTCCCAGTACAAATTTTATTTAGCTTTTGAA 430

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
```

```
          |||||||| ||| | || |||||||| || || || |    | |||||  |      |  |
Db    431 AATTCAATTCACGAAGACTACATCACTGAGAAACTCTGGCATAATGCCCCTCTCATCGTGG 490

Qy    628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
          || |||||||||||| |||| || ||   |    |||||||||  || | | ||  | ||
Db    491 GCTCTACCTGTTGTTTTGGGGCCCTCGCGTTCTAACTATGAGCGTTTTTTACCCCCCGAG 550

Qy    688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 747
          |||||||| || || || || | |    || | |  |  || ||  | ||||| || ||
Db    551 TCATTCATCCACGTAGATGACTTTCCAACTGCTAAGGGACTTGCTGATTATCTAAACGAG 610

Qy    748 GTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 788
          | || ||| |  ||||   |||| |     | || |||||
Db    611 CTTGATAAAGATGATAATAAATACATGCAACATTTCAACTG 651


RESULT 13
BJ050300/c
LOCUS       BJ050300                666 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  BJ050300 NIBB Mochii normalized Xenopus neurula library Xenopus
            laevis cDNA clone XL023o15 3', mRNA sequence.
ACCESSION   BJ050300
VERSION     BJ050300.1  GI:17392988
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL023o15"
                     /clone_lib="NIBB Mochii normalized Xenopus neurula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 15"
BASE COUNT      178 a    116 c    157 g    215 t
ORIGIN

  Query Match             16.6%;  Score 151.6;  DB 10;  Length 666;
  Best Local Similarity   54.5%;  Pred. No. 4e-31;
  Matches  348;  Conservative    0;  Mismatches  284;  Indels    6;  Gaps    2;

Qy    251 CAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACC 310
          |||  |||| |||| | |        || ||| |  |  | || |||||  || ||||||
Db    666 CAAGCCACAATCCCAACACAGCGTTTATGGAGAAACTCATCAATCTGACCATGTCTTACC 607

Qy    311 GCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCG 370
          |   ||| ||||||||       ||||||||||   |
Db    606 GTGCTGACTCAGATATTTTTACTCCTTATGGCTGGCTAGAAAGGCATGATGGAAAAGAAA 547

Qy    371 TGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTG 430
            ||   ||||  |||   ||| | ||  |||||| || ||||||||||||
Db    546 ACTTCACTATGCCCGCCAAGACAAAACTAGTAGCCTGGGTAGTCAGTAACTGGAACCCAA 487

Qy    431 AGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACG 490
          |       ||| |||| |||||  | ||| |||    |  | |||   |  ||| ||| |
Db    486 ACTCCAGAAGAATCAAATATTATGAGGAGTTAAAGCCACACCTTGCTGTTGATATCTATG 427

Qy    491 GGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTA 550
          | | |      |  |      |   |      ||| |    |||   |  | ||    |  |
Db    426 GCCTACAGCATTTACCGCTTCCTAGACCAAAACA---CATTGAAATTCTTTCCCAGTACA 370

Qy    551 AATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT--- 607
          |||||||| |  | ||||||||||||||| ||| | || ||||||||| || || || |
Db    369 AATTTTATTTAGCTTTTGAAAATTCAATTCACGAAGACTACATCACTGAGAAACTCTGGC 310

Qy    608 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 667
          | |||||   |    |  |  || |||||||||| |||| || ||  |    |||||||
Db    309 ATAATGCCCTCTCATCGTGGGCTCTACCTGTTGTTTTGGGGCCCTCGCGTTCTAACTATG 250

Qy    668 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 727
          ||  || | | ||  | || |||||||| || || || || | |    || | |  |  |
Db    249 AGCGTTTTTTACCCCCCGAGTCATTCATCCACGTAGATGACTTTCCAACTGCTAAGGGAC 190

Qy    728 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 787
          | ||  | ||||| || ||  | || ||| |  ||||   |||| |     | || ||||
Db    189 TTGCTGATTATCTAAACGAGCTTGATAAAGATGATAATAAATACATGCAACATTTCAACT 130

Qy    788 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 847
          |||||||  || | |     |       ||       |||| |  | |||   ||    | |
Db    129 GGAGGAAACATCTAAAACCAGTAGCACCTGAGAGTTGGAACCCCCATTACTGCAAAGTAT 70

Qy    848 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTT 885
          || |   | | |||       |     |||| | || ||
Db     69 GCAAAGCTATAAAAGACGCACCGTCTTATAGGACTTTT 32


RESULT 14
CNS034XF/c
LOCUS       CNS034XF                711 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            212J18 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL228012
VERSION     AL228012.1  GI:7886987
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 711)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 711)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="212J18"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG212DE09SP1-end :
```

```
                    PUC-Ori"
BASE COUNT      185 a    153 c    168 g    203 t      2 others
ORIGIN

  Query Match             16.5%;  Score 151.2;  DB 12;  Length 711;
  Best Local Similarity   55.0%;  Pred. No. 5.3e-31;
  Matches  360;  Conservative    1;  Mismatches  284;  Indels    9;  Gaps    3;

Qy     170 ACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGA 229
           ||||    ||| ||||| |||| || ||   | |  |  |||||   || |||| ||||
Db     705 ACATACAGTGGAATCTGGCAAACTTTCCAGTGGAGCCGCGGCCATATTTTCAGAGATGGG 646

Qy     230 TTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGT 289
           ||||| |  || | ||| | ||||   ||||   |   |    |||  | ||     | |
Db     645 TTTGGTTCTATCTAGAACCTCCAATGAACACATACAGGATACCTGGTCTAGAAACACTCT 586

Qy     290 TTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGA 349
           | ||| ||||||||   ||| |    || || || ||    |     ||  . |    |
Db     585 TCAACATGACTCTGCACTACAGGAAAGACTCCGACATTGTGGCCAGGTACCCTCTTACCA 526

Qy     350 CGGTAAGCACAAATCCCTTCGTGTTTGAAG---TGCCAAGCAAAGAGAAATTGGTGTGCT 406
           |     ||  |     | |     ||   ||||    ||| | |||   ||||| |
Db     525 TCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGAGAAAAACAAAACTGTGTGTT 466

Qy     407 GGGTTGTGAG---TAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           || | || ||   |||  | |          ||| |:|    | | |||   ||| |||
Db     465 GGATCGTCAGCAATAATCGTACATTCACTGGTGCAAACAGCAATTTTATCGTGAACTAT 406

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCAT---TTGGAGAATATGTCAATGATA 520
           | |    ||| | ||  ||   || ||  | || |     || |      |  |   |
Db     405 CCCAGCATATTTACATAAATGTTTATGGTTACGCCTACACAGGGGTCAGGTTGGAGATTG 346

Qy     521 AAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCC 580
           |  | |   | |  ||| |  ||   ||||| ||||| || |||||||| || ||||||
Db     345 ATCAATATTTCCAGACCGTTGCTAGCTGTAAGTTTTACCTGTCCTTTGAGAACTCAATCT 286

Qy     581 ACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           |||| ||||||||||||| ||| |  ||    |  ||   |  ||  | || ||||| |
Db     285 ACAAAGATTACATCACGGAGAAGGTCAACGGACCCCTTGCAGTAGGAACCGTTCCTGTAG 226

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||||  | || |  ||||| || |  |   | ||  |||| || || ||| |
Db     225 TTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTTCCCTTCAGACTCCTTTATTAACA 166

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACA 760
           | || |||| |     ||| |  |  || ||| || |||||       | |||||  |
Db     165 TAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCATCTGCTCTCCTTGGACAAGGATG 106

Qy     761 ATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTC 814
           |  ||  |||| |  | |||||| | |||||||| | ||||  || | || ||
Db     105 AGGAGGCATACATGCGCTACTTTGAGTGGAGGAAGTACTTCAGCGTCATTCCTC 52


RESULT 15
B41073/c
LOCUS       B41073                   501 bp    DNA     linear   GSS 18-OCT-1997
DEFINITION  HS-1052-B2-G01-MF.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 774 Col=2 Row=N, DNA sequence.
ACCESSION   B41073
VERSION     B41073.1  GI:2545325
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 774  row: N  column: 2
            Class: BAC ends
            High quality sequence stop: 501.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=CT 774 Col=2 Row=N"
                     /clone_lib="CIT Human Genomic Sperm Library C"
                     /sex="M"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      179 a     95 c     83 g    143 t      1 others
ORIGIN

  Query Match             16.5%;  Score 151;  DB 12;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-31;
  Matches  151;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     765 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 824
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 442

Qy     825 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 884
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 382

Qy     885 TGGTAATTTAGAGAAATGGTTTTGGAATTAA 915
           ||||||||||||||||||||||||||||||
Db     381 TGGTAATTTAGAGAAATGGTTTTGGAATTAA 351


Search completed: October  6, 2002, 07:15:58
Job time : 772.105 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 18.6694 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-4_COPY_1454_2368
Perfect score: 915
Sequence: 1 accaaaactgattatttttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	13.9	1814	2	US-08-483-151-1	Sequence 1, Appli
2	127	13.9	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	113.4	12.4	1654	1	US-07-914-281-13	Sequence 13, Appl
4	113.4	12.4	1654	1	US-08-393-246-13	Sequence 13, Appl
5	113.4	12.4	1654	1	US-08-525-058A-13	Sequence 13, Appl
6	113.4	12.4	1654	2	US-08-696-731-13	Sequence 13, Appl
7	113.4	12.4	1654	4	US-09-042-531-13	Sequence 13, Appl
8	112.2	12.3	1086	1	US-07-914-281-12	Sequence 12, Appl
9	112.2	12.3	1086	1	US-08-393-246-12	Sequence 12, Appl
10	112.2	12.3	1086	1	US-08-525-058A-12	Sequence 12, Appl
11	112.2	12.3	1086	2	US-08-696-731-12	Sequence 12, Appl
12	112.2	12.3	1086	4	US-09-042-531-12	Sequence 12, Appl
13	112.2	12.3	2042	3	US-09-063-237-2	Sequence 2, Appli
14	112.2	12.3	2043	1	US-07-914-281-1	Sequence 1, Appli
15	112.2	12.3	2043	1	US-08-393-246-1	Sequence 1, Appli
16	112.2	12.3	2043	1	US-08-273-411-4	Sequence 4, Appli
17	112.2	12.3	2043	1	US-08-525-058A-1	Sequence 1, Appli
18	112.2	12.3	2043	2	US-08-696-731-1	Sequence 1, Appli
19	112.2	12.3	2043	4	US-09-042-531-1	Sequence 1, Appli
20	112.2	12.3	2043	5	PCT-US91-00899-1	Sequence 1, Appli
21	111.8	12.2	1316	1	US-07-914-281-10	Sequence 10, Appl
22	111.8	12.2	1316	1	US-08-393-246-10	Sequence 10, Appl
23	111.8	12.2	1316	1	US-08-525-058A-10	Sequence 10, Appl
24	111.8	12.2	1316	2	US-08-696-731-10	Sequence 10, Appl
25	111.8	12.2	1316	4	US-09-042-531-10	Sequence 10, Appl
26	100.8	11.0	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	100.8	11.0	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	100.8	11.0	1488	1	US-07-914-281-9	Sequence 9, Appli
29	100.8	11.0	1488	1	US-08-393-246-9	Sequence 9, Appli
30	100.8	11.0	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	100.8	11.0	1488	2	US-08-696-731-9	Sequence 9, Appli
32	100.8	11.0	1488	4	US-09-042-531-9	Sequence 9, Appli
33	100.8	11.0	2175	4	US-08-482-073-9	Sequence 9, Appli
34	100.8	11.0	2861	4	US-08-482-073-10	Sequence 10, Appl
35	100.8	11.0	3647	1	US-07-914-281-7	Sequence 7, Appli
36	100.8	11.0	3647	1	US-08-393-246-7	Sequence 7, Appli
37	100.8	11.0	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	100.8	11.0	3647	2	US-08-696-731-7	Sequence 7, Appli
39	100.8	11.0	3647	4	US-09-042-531-7	Sequence 7, Appli
40	100.8	11.0	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	97.6	10.7	2134	2	US-08-483-151-3	Sequence 3, Appli
42	97.6	10.7	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	37.2	4.1	1728	1	US-08-403-866-12	Sequence 12, Appl
c 44	37.2	4.1	2231	1	US-08-403-866-14	Sequence 14, Appl
c 45	37.2	4.1	12720	1	US-08-403-866-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
;  GENERAL INFORMATION:
;    APPLICANT:  Seed, Brian
;    APPLICANT:  Holgersson, Jan
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-151-1

  Query Match             13.9%;  Score 127;  DB 2;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 9.7e-29;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy    104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
          | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |          |   |     |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
             ||| | |||       ||||||| || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
             | | || |||   |   ||||  || || ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
             ||||  |     | |    ||   |   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
             |||||| | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
             |       |  |  |  ||   | ||| |  ||    |||       |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
             | || | || ||  |  |     |      || || ||  |||||| || ||| ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
             ||| |||||||| || ||| | |    ||||||  |   ||||||  |||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
             ||||||||  || ||| | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           |||| || |  | |||| || |||| || ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175.


RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;    APPLICANT:  The General Hospital Corporation
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             13.9%;  Score 127;  DB 5;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 9.7e-29;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
           | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |          |   |     |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
             ||| | |||       ||||||| || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
             | | || |||   |   ||||  || || ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
             ||||  |     | |    ||   |   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
             |||||| | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
             |       |  |  |  ||   | ||| |  ||    |||       |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
             | || | || ||  |  |     |      || || ||  |||||| || ||| ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
             ||| |||||||| || ||| | |    ||||||  |   ||||||  |||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
             ||||||||  || ||| | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           |||| || |  | |||| || |||| || ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 3
US-07-914-281-13
; Sequence 13, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-13

  Query Match            12.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 1.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| ||||| |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
            |      ||     |  ||| |  ||||  | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||          | | | |      |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||      |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           |      |     |     |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    |||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |   |    |||| || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           |||
Db    1095 GAG 1097
```

```
RESULT 4
US-08-393-246-13
; Sequence 13, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-13

  Query Match            12.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 1.2e-24;
```

```
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |      ||      |  ||| |  ||||  | |||||     ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||          | | | |     |   | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
             |      |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |  |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| |  ||  | ||||| || ||||| || |    |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| ||||  | |||||  || |       |||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           |||
Db    1095 GAG 1097


RESULT 5
US-08-525-058A-13
; Sequence 13, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-13

  Query Match           12.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity  51.2%;  Pred. No. 1.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |      ||      |  ||| |  ||||  | |||||     ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||          | | | |     |   | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
             |      |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914
```

```
Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| ||||  | |||||  || |      |||||| || |||||| |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           ||
Db    1095 GAG 1097
```

RESULT 6
US-08-696-731-13
; Sequence 13, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,731
; FILING DATE: 14-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/393,246
; FILING DATE:
; APPLICATION NUMBER: US 08/220,433
; FILING DATE: 30-MAR-1994
; APPLICATION NUMBER: US 07/914,281
; FILING DATE: 20-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-696-731-13

Query Match 12.4%; Score 113.4; DB 2; Length 1654;
Best Local Similarity 51.2%; Pred. No. 1.2e-24;
Matches 401; Conservative 0; Mismatches 361; Indels 21; Gaps 5;

```
Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
           ||   ||      | ||| |  |||| | |||||     ||||      |  | |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||          | | | |     |    | |    | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||      |  ||||| |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||          |  |  |  |    ||||  |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           |     |     |    ||  ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
           ||| ||| ||||  | |||||  || |      |||||| || |||||| |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           ||
Db    1095 GAG 1097
```

RESULT 7
US-09-042-531-13
; Sequence 13, Application US/09042531
; Patent No. 6268193
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor

```
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/042,531
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1654 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-13

  Query Match             12.4%;  Score 113.4;  DB 4;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||    |  || | |  ||  |   |||| ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
            |     ||     |  ||| |  |||| | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |   |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||          | | | |      |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  | |  |||     |  |||||  |||||  |||   ||||||  ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           |||||  ||    || ||| |    ||           |  |  |  |    ||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           |     |    |    |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || ||||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  || || || || ||||| ||    ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | |||  | ||  | |||||  || ||||| || |   |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
           |||  |||  ||||  | |||||  || |       |||||  || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           ||
Db    1095 GAG 1097


RESULT 8
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;
```

```
Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  |||  | ||  |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |   |  |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||     ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          |||||  || |||||  |||  |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |       |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 9
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  |||  | ||  |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |   |  |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||     || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||     ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          |||||  || |||||  || |   |    ||||  || || ||   ||| ||| ||||  |
```

```
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |      |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 10
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |    | |    |     |          | |  |  |  |   |||

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||          |  |  |  |     |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || || || |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||    || ||||||||| || ||||| |     || ||  |   ||| |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||    ||    |||||  |||  |   | |||  | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |    |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |      |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 11
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
```

```
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             12.3%;  Score 112.2;  DB 2;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |   |  | ||    | || |       | |||||      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |    |  |    |              | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
            |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
            |    ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || ||||| ||    ||    ||||| |||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 12
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match             12.3%;  Score 112.2;  DB 4;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |   |  | ||    | || |       | |||||      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |    |  |    |              | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
            |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
            |    ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690
```

```
Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
           || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db     691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
           ||  | |||   || |||||||| || ||||| |     || ||  |  |||  |  |
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
           || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
           ||||| || ||||| || |    |    |||| || || ||   ||| ||| ||||  |
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
           |||||  || |       ||||| || ||||||  |||| || ||
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 13
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             12.3%;  Score 112.2;  DB 3;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 3e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
           |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
           |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
           |  || || || ||  || | |||||| | || || || |||    ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
           |   ||  |    |  |    |     |          | |  |  |  |   |||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
           |  |||||  |||||  |||   |||||||| || ||    ||||| ||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
           |   ||          |  |  |  |  ||||| |  |                    |
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
           || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
           ||  | |||   || |||||||| || ||||| |     || ||  |  |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
           || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
           ||||| || ||||| || |    |    |||| || || ||   ||| ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
           |||||  || |       ||||| || ||||||  |||| || ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match            12.3%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity  51.6%;  Pred. No. 3e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || |||| |   |   |  | ||    | || |      | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |   |   |     |          | |  |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||           |  |  |  |    |||||  |   |               |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||     ||    ||||| ||||  |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |      |||||  || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 15
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match            12.3%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity  51.6%;  Pred. No. 3e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || |||| |   |   |  | ||    | || |      | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
```

```
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453

Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513

Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573

Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630

Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690

Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750

Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795

Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:21:30
Job time : 28.6694 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 3353.32 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-5
Perfect score: 2676
Sequence: 1 gcgcgcgcggcgcagcagct.....ccaaataaacctaccattac 2676

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2676	100.0	3019	9	AB023021	AB023021 Homo sapi
2	2423.6	90.6	2501	9	HSA238701	AJ238701 Homo sapi
3	2380.4	89.0	80247	9	AL512406	AL512406 Human DNA
4	1206.8	45.1	2139	10	AB015426	AB015426 Mus muscu
5	1184.8	44.3	2156	10	AB049819	AB049819 Rattus no
6	1147.6	42.9	1705	10	AF230460	AF230460 Cricetulu
7	958.2	35.8	1128	10	AF345993	AF345993 Rattus no
8	820.4	30.7	1707	5	AB035906	AB035906 Gallus ga
9	676.8	25.3	1080	5	AB035905	AB035905 Xenopus l
10	639	23.9	71396	2	AC100674	AC100674 Mus muscu
c 11	343.4	12.8	71396	2	AC100674	AC100674 Mus muscu
12	279.6	10.4	1567	5	AB023627	AB023627 Danio rer
13	234.6	8.8	1832	5	AB023628	AB023628 Danio rer
14	222.6	8.3	1092	5	AB035907	AB035907 Xenopus l
15	220.2	8.2	1264	5	AF288369	AF288369 Gallus ga
c 16	197.4	7.4	120514	9	HS84D21	AL034348 Human DNA
17	197.4	7.4	191190	2	AL356303	AL356303 Homo sapi
18	197	7.4	1180	9	BC001879	BC001879 Homo sapi
19	176.2	6.6	2657	5	GGU73678	U73678 Gallus gall
20	145.2	5.4	1429	10	CGU78737	U78737 Cricetulus
21	145.2	5.4	1462	10	AF090450	AF090450 Cricetulu
22	143.6	5.4	1718	10	AF090449	AF090449 Cricetulu
23	137.2	5.1	1044	5	AB035908	AB035908 Gallus ga
24	132	4.9	2293	4	BTA134FTG	X87810 B.taurus al
25	132	4.9	12607	4	BTA132772	AJ132772 Bos tauru
26	127.6	4.8	1158	10	AB039126	AB039126 Mus muscu
27	127	4.7	1487	3	AF016899	AF016899 Schistoso
28	127	4.7	1814	6	AR028686	AR028686 Sequence
29	127	4.7	3594	10	MMU45980	U45980 Mus musculu
c 30	127	4.7	117993	2	AC091284	AC091284 Mus muscu
31	125.6	4.7	1613	4	BTA132776	AJ132776 Bos tauru
32	125.6	4.7	1619	4	BTA132775	AJ132775 Bos tauru
33	125.6	4.7	1644	4	BTA132774	AJ132774 Bos tauru
34	125.6	4.7	1873	4	BTA132773	AJ132773 Bos tauru
35	124.6	4.7	1119	9	AF345881	AF345881 Macaca mu
36	122.8	4.6	1170	10	AB039127	AB039127 Mus muscu
37	122.8	4.6	1170	10	AB039131	AB039131 Mus muscu
38	122.8	4.6	1615	10	MMU33457	U33457 Mus musculu
39	122	4.6	1113	9	AF345884	AF345884 Saimiri s
40	121.2	4.5	1170	10	AB039124	AB039124 Mus muscu
41	121.2	4.5	1170	10	AB039125	AB039125 Mus muscu
42	121.2	4.5	1170	10	AB039128	AB039128 Mus muscu
43	121.2	4.5	1170	10	AB039129	AB039129 Mus muscu
44	121.2	4.5	1170	10	AB039130	AB039130 Mus muscu
45	121.2	4.5	1170	10	AB039132	AB039132 Mus muscu

ALIGNMENTS

RESULT 1
AB023021
LOCUS AB023021 3019 bp mRNA linear PRI 23-JUN-1999
DEFINITION Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX, complete cds.
ACCESSION AB023021
VERSION AB023021.1 GI:5139692
KEYWORDS alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S., Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and Narimatsu,H.
TITLE Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX
JOURNAL FEBS Lett. 453, 237-242 (1999)
REFERENCE 2 (bases 1 to 3019)
AUTHORS Kaneko,M., Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
ORIGIN

  Query Match           100.0%;  Score 2676;  DB 9;  Length 3019;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2676;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCG 60

Qy      61 CCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGCCGCCAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGCCGCCAC 120

Qy     121 CGCCGCCTGGCCCTGCCTGCCTCCTGCGCCCGCGCAGCCCTCGCGAGCGCCCCGGATGGCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGCCGCCTGGCCCTGCCTGCCTCCTGCGCCCGCGCAGCCCTCGCGAGCGCCCCGGATGGCG 180

Qy     181 CTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGT 240

Qy     241 AGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACA 300

Qy     301 TCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTC 360

Qy     361 ATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAA 420

Qy     421 TCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAAT 480

Qy     481 GAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 540

Qy     541 CAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAA 600

Qy     601 TCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     601 TCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAG 660

Qy     661 CAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACT 720

Qy     721 CCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCA 780

Qy     781 GATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTG 840

Qy     841 CCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGA 900

Qy     901 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 960

Qy     961 GGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTT 1020

Qy    1021 TCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTG 1080

Qy    1081 GCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCA 1140

Qy    1141 GCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 1200

Qy    1201 AAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 1260

Qy    1261 ACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAA 1320

Qy    1321 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTT 1380

Qy    1381 TCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAG 1440

Qy    1441 AAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 AAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAA 1500

Qy    1501 CGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTT 1560

Qy    1561 TAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCA 1620

Qy    1621 TTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGA 1680

Qy    1681 TTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 TTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTA 1740
```

```
Qy   1741 AAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAAC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAAC 1800

Qy   1801 AGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTG 1860

Qy   1861 TATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAA 1920

Qy   1921 TTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTA 1980

Qy   1981 GACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATT 2040

Qy   2041 AAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTTCAGCACCTGGGAAGTAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTTCAGCACCTGGGAAGTAA 2100

Qy   2101 TCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTTCTGCTACTGATGACACTCATTGT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTTCTGCTACTGATGACACTCATTGT 2160

Qy   2161 CATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAATGATACCTATAAATATATTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAATGATACCTATAAATATATTT 2220

Qy   2221 ATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGA 2280

Qy   2281 AGCCCTAATATTTAAAATGGCCTTATTTTACCATATGGATATAAGATTTGGCTCATAATG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AGCCCTAATATTTAAAATGGCCTTATTTTACCATATGGATATAAGATTTGGCTCATAATG 2340

Qy   2341 ATGAGCCCTATCATTTGATTTGAGTTCTATCATTTAAGAGAGCCTAAATAAAATTATCAT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 ATGAGCCCTATCATTTGATTTGAGTTCTATCATTTAAGAGAGCCTAAATAAAATTATCAT 2400

Qy   2401 CAAGGTATTAAATATAAGACGTTAAATATAATAAAGTGGGGATATATAGAAAACACACAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CAAGGTATTAAATATAAGACGTTAAATATAATAAAGTGGGGATATATAGAAAACACACAG 2460

Qy   2461 TGTTAGCACAGAGTAAGATCTCAATGCACATTTGTTGGATGAATAAATAAATGCAATTGA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TGTTAGCACAGAGTAAGATCTCAATGCACATTTGTTGGATGAATAAATAAATGCAATTGA 2520

Qy   2521 ATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAGTTCTACTTTAGAAGTACTAATTGG 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 ATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAGTTCTACTTTAGAAGTACTAATTGG 2580

Qy   2581 AGATGACTTTTATATCCCATTTTGGTAATTATTCATACATAGCACATATGACCATGATGT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGATGACTTTTATATCCCATTTTGGTAATTATTCATACATAGCACATATGACCATGATGT 2640

Qy   2641 TCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 2676
          ||||||||||||||||||||||||||||||||||
Db   2641 TCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 2676
```

```
RESULT 2
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match             90.6%;  Score 2423.6;  DB 9;  Length 2501;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 2470;  Conservative    0;  Mismatches    4;  Indels    4;  Gaps    4;

Qy     199 TTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTAC 258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 TTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTAC 68

Qy     259 GTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATT 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 GTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATT 128

Qy     319 CTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATT 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 CTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATT 188

Qy     379 TACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTG 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 TACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTG 248
```

```
Qy    439 AAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTG 498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 AAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTG 308

Qy    499 TGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATC 558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 TGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATC 368

Qy    559 CAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATC 618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATC 428

Qy    619 CATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTC 678
          |||||||||||||||| |||||||||| || ||||||||||||||||||||||||||||||
Db    429 CATCACCGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAAGCTAGGCCACCCTTC 488

Qy    679 CAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATT 738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATT 548

Qy    739 GAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTAT 798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 GAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTAT 608

Qy    799 GGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTG 858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 GGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTG 668

Qy    859 GTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAG 918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAG 728

Qy    919 CTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGAT 978
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 CTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGAT 788

Qy    979 AAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATC 1038
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATC 848

Qy   1039 CACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTT 1098
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 CACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTT 908

Qy   1099 GTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCAT 1158
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 GTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCAT 968

Qy   1159 GTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAAC 1218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAAC 1028

Qy   1219 AATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGA 1278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 AATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGA 1088

Qy   1279 TTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAG 1338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 TTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAG 1148

Qy   1339 TCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTG 1398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 TCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTG 1208

Qy   1399 ATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTA 1458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 ATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTA 1268

Qy   1459 CTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGA 1518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 CTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGA 1328
```

```
Qy   1519 GATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATA 1578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATA 1388

Qy   1579 CCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTT 1638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 CCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTT 1448

Qy   1639 TCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGA 1698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 TCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGA 1508

Qy   1699 TCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGAAAAATTTTCA 1758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509 TCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGAAAAATTTTCA 1568

Qy   1759 CTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAACATGTTGC 1818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1569 CTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAACATGTTGC 1628

Qy   1819 GATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTGTATAACTGTCTCTATTTG 1878
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1629 GATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTGTATAACTGTCTCTA-TTG 1687

Qy   1879 ATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAATTATTCATGGAGTGAATA 1938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1688 ATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAATTATTCATGGAGTGAATA 1747

Qy   1939 AGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTAGACTTCTCATTTATTTTC 1998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1748 AGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTAGACTTCTCATTTATTTTC 1807

Qy   1999 ATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATTAAATTATTCAAGTATTTT 2058
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1808 ATTAAGCTGATTTGCAGCTACTTATTCTCATGGCCTTAAATTAAATTATTCAAGTATTTT 1867

Qy   2059 TAAATATCCAATTTGTTGTGATTTTCAGCACCTGGGAAGTAATCCCAATAATACTTTAGA 2118
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1868 TAAATATCCAATTTGTTGTGATTTTCAGCACCT-GGAAGTAATCCCAATAATACTTTAGA 1926

Qy   2119 AAATCTAAGACAGTTCTTTCTGCTACTGATGACACTCATTGTCATAATAAAACAAATAAT 2178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1927 AAATCTAAGACAGTTCTTTCTGCTACTGATGACACTCATTGTCATAATAAAACAAATAAT 1986

Qy   2179 TTCCTCAAATAACAAAGAAAAATGATACCTATAAATATATTTATAAATGGTGTCATTTAT 2238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1987 TTCCTCAAATAACAAAGAAAAATGATACCTATAAATATATTTATAAATGGTGTCATTTAT 2046

Qy   2239 GAACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAAAAT 2298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2047 GAACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAAAAT 2106

Qy   2299 GGCCTTATTTTACCATATGGATATAAGATTTGGCTCATAATGATGAGCCCTATCATTTGA 2358
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   2107 GGCCTTATTTTACCATATGGATATAAGATTTGGCTCATAATGATGAG-CCTATCATTTGA 2165

Qy   2359 TTTGAGTTCTATCATTTAAGAGAGCCTAAATAAAATTATCATCAAGGTATTAAATATAAG 2418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2166 TTTGAGTTCTATCATTTAAGAGAGCCTAAATAAAATTATCATCAAGGTATTAAATATAAG 2225

Qy   2419 ACGTTAAATATAATAAAGTGGGGATATATAGAAAACACACAGTGTTAGCACAGAGTAAGA 2478
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   2226 ACGTTAAATATAATAAAGTGGGGATATATAG-AAACACACAGTGTTAGCACAGAGTAAGA 2284

Qy   2479 TCTCAATGCACATTTGTTGGATGAATAAATAAATGCAATTGAATTCCCAGAAAAATGATT 2538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2285 TCTCAATGCACATTTGTTGGATGAATAAATAAATGCAATTGAATTCCCAGAAAAATGATT 2344

Qy   2539 GTTTCAAGGAAGTGACAGTTCTACTTTAGAAGTACTAATTGGAGATGACTTTTATATCCC 2598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2345 GTTTCAAGGAAGTGACAGTTCTACTTTAGAAGTACTAATTGGAGATGACTTTTATATCCC 2404

Qy   2599 ATTTTGGTAATTATTCATACATAGCACATATGACCATGATGTTCAGGGCTTTATAGAACC 2658
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2405 ATTTTGGTAATTATTCATACATAGCACATATGACCATGATGTTCAGGGCTTTATAGAACC 2464

Qy    2659 AAATAAACCTACCATTAC 2676
          ||||||||||||||||||
Db    2465 AAATAAACCTACCATTAC 2482


RESULT 3
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-504J9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-381A2 is at 80148 in this sequence.
            The true right end of clone RP11-77B15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .80247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-504J9"
                     /clone_lib="RPCI-11.2"
     repeat_region   184. .1422
                     /note="L2 repeat: matches 875. .2419 of consensus"
     repeat_region   1744. .1834
                     /note="MIR repeat: matches 166. .258 of consensus"
     repeat_region   2950. .2977
                     /note="14 copies 2 mer ca 100% conserved"
     repeat_region   3411. .3669
                     /note="LTR16C repeat: matches 93. .364 of consensus"
     repeat_region   4372. .4704
                     /note="MLT1A1 repeat: matches 5. .365 of consensus"
     misc_feature    complement(7821. .8406)
                     /note="match: GSS: Em:AQ550748"
     repeat_region   8126. .8184
                     /note="MER58B repeat: matches 272. .332 of consensus"
     repeat_region   8185. .8696
                     /note="MLT2D repeat: matches 1. .513 of consensus"
     repeat_region   8779. .8817
                     /note="MLT2D repeat: matches 504. .542 of consensus"
     repeat_region   8819. .8974
                     /note="MER58A repeat: matches 6. .162 of consensus"
     misc_feature    complement(9443. .9891)
                     /note="match: GSS: Em:B90633"
     misc_feature    complement(9452. .9883)
                     /note="match: GSS: Em:AQ119695"
     repeat_region   9782. .9894
                     /note="MIR repeat: matches 92. .205 of consensus"
     repeat_region   10089. .10212
                     /note="MIR repeat: matches 135. .262 of consensus"
     misc_feature    complement(10093. .10671)
                     /note="match: GSS: Em:AQ237061"
     repeat_region   10223. .10282
                     /note="MIR repeat: matches 203. .262 of consensus"
     misc_feature    complement(10315. .10631)
                     /note="match: GSS: Em:AQ200259"
     misc_feature    10805. .11186
                     /note="match: GSS: Em:AQ072353"
     repeat_region   11657. .11844
                     /note="L2 repeat: matches 2546. .2750 of consensus"
     misc_feature    complement(11659. .12084)
                     /note="match: GSS: Em:AQ518361"
     misc_feature    complement(11674. .11987)
                     /note="match: GSS: Em:AQ094678"
     misc_feature    12083. .12588
                     /note="match: GSS: Em:AQ691235"
     repeat_region   12596. .12810
                     /note="MIR repeat: matches 23. .248 of consensus"
     repeat_region   12920. .12986
                     /note="L1MA5 repeat: matches 6228. .6295 of consensus"
     repeat_region   12987. .13169
                     /note="L1MA5 repeat: matches 5852. .6043 of consensus"
     repeat_region   13168. .13280
                     /note="L1MA8 repeat: matches 5710. .5832 of consensus"
     repeat_region   13851. .17218
                     /note="L1PB1 repeat: matches 2787. .6155 of consensus"
     repeat_region   18903. .18953
                     /note="MLT2CB repeat: matches 439. .501 of consensus"
     repeat_region   18984. .19433
                     /note="MLT2CB repeat: matches 1. .460 of consensus"
     misc_feature    complement(19108. .19690)
                     /note="match: GSS: Em:AQ563935"
     repeat_region   20683. .20801
                     /note="AluJo/FRAM repeat: matches 169. .288 of consensus"
     repeat_region   20802. .20862
                     /note="AluJ/FLAM repeat: matches 3. .66 of consensus"
     repeat_region   20912. .21033
                     /note="MIR repeat: matches 114. .237 of consensus"
     repeat_region   21047. .21128
                     /note="41 copies 2 mer ta 63% conserved"
     repeat_region   21676. .21859
                     /note="L2 repeat: matches 2130. .2308 of consensus"
     repeat_region   21890. .23612
                     /note="L1P repeat: matches 841. .2570 of consensus"
     repeat_region   23601. .26609
                     /note="L1PA5 repeat: matches 3107. .6145 of consensus"
     repeat_region   26620. .26789
                     /note="MER5A repeat: matches 1. .174 of consensus"
     repeat_region   28521. .28627
                     /note="MER39 repeat: matches 5. .111 of consensus"
     repeat_region   28656. .28837
                     /note="MER39b repeat: matches 54. .235 of consensus"
     repeat_region   28838. .29142
```

```
                /note="AluSc repeat: matches 4. .308 of consensus"
     repeat_region   29143. .29494
                /note="MER39b repeat: matches 235. .549 of consensus"
     repeat_region   30529. .31344
                /note="L1PA3 repeat: matches 5331. .6146 of consensus"
     repeat_region   32234. .32340
                /note="L2 repeat: matches 2627. .2750 of consensus"
     repeat_region   32399. .32578
                /note="MIR repeat: matches 12. .203 of consensus"
     repeat_region   32816. .32979
                /note="MIR repeat: matches 91. .259 of consensus"
     misc_feature    complement(33006. .33450)
                /note="match: GSS: Em:AQ261227"
     misc_feature    complement(33058. .33444)
                /note="match: GSS: Em:AQ040453"
     repeat_region   33312. .33596
                /note="L1M4 repeat: matches 3473. .3773 of consensus"
     repeat_region   34856. .34915
                /note="30 copies 2 mer aa 68% conserved"
     misc_feature    complement(34987. .35547)
                /note="match: GSS: Em:AQ321598
                match: STS: Em:G55571"
     misc_feature    complement(36488. .36936)
                /note="match: GSS: Em:B41073"
     misc_feature    complement(36654. .37090)
                /note="match: GSS: Em:AQ772584"
     misc_feature    37102. .37591
                /note="match: GSS: Em:AQ489567"
     misc_feature    complement(37581. .38221)
                /note="match: GSS: Em:AQ418233"
     repeat_region   37720. .37772
                /note="L2 repeat: matches 2657. .2710 of consensus"
     misc_feature    complement(37778. .38215)
                /note="match: GSS: Em:AQ819773"
     misc_feature    38218. .38732
                /note="match: GSS: Em:AQ599868"
     repeat_region   38861. .39164
                /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   39612. .39653
                /note="21 copies 2 mer ac 90% conserved"
     repeat_region   40854. .40909
                /note="L2 repeat: matches 2655. .2710 of consensus"
     repeat_region   42169. .42270
                /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   42411. .42548
                /note="MER5B repeat: matches 21. .181 of consensus"
     repeat_region   42780. .42938
                /note="MER5B repeat: matches 17. .177 of consensus"
     repeat_region   43229. .43326
                /note="MIR repeat: matches 40. .148 of consensus"
     repeat_region   45383. .45428
                /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                /note="match: STS: Em:G51995
                match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
                /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match             89.0%;  Score 2380.4;  DB 9;  Length 80247;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 2384;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     287 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35551 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 35610

Qy     347 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35611 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 35670

Qy     407 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35671 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 35730

Qy     467 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35731 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 35790

Qy     527 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 586
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35791 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 35850

Qy     587 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35851 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 35910

Qy     647 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 706
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35911 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 35970

Qy     707 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35971 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 36030

Qy     767 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36031 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 36090

Qy     827 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36091 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 36150

Qy     887 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36151 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 36210

Qy     947 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 1006
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   36211 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTACTT 36270

Qy    1007 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36271 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 36330

Qy    1067 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36331 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 36390

Qy    1127 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36391 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 36450

Qy    1187 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36451 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 36510

Qy    1247 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36511 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 36570

Qy    1307 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36571 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 36630

Qy    1367 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 1426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36631 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 36690

Qy    1427 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 1486
```

```
Db  36691 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 36750

Qy   1487 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 1546
Db  36751 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 36810

Qy   1547 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 1606
Db  36811 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 36870

Qy   1607 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 1666
Db  36871 ATTCTTCATTATCATTTGTAAACATTGTTTTTTCACATTTTTGTAGTTGTCCATAATGTA 36930

Qy   1667 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 1726
Db  36931 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 36990

Qy   1727 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 1786
Db  36991 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 37050

Qy   1787 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGACTCCTA 1846
Db  37051 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGACTCCGA 37110

Qy   1847 AGATGAATGAAGTGTATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACAT 1906
Db  37111 AGATGAATGAAGTGTATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACAT 37170

Qy   1907 TTGTGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTAT 1966
Db  37171 TTGTGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTAT 37230

Qy   1967 TCAGGAAGCTATTAGACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCT 2026
Db  37231 TCGGGAAGCTATTAGACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCT 37290

Qy   2027 CATGGTCTTAAATTAAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTTCAG 2086
Db  37291 CATGGTCTTAAATTAAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTTCAG 37350

Qy   2087 CACCTGGGAAGTAATCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTTCTGCTACTG 2146
Db  37351 CACCTGGGAAGTAATCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTTCTGCTACTG 37410

Qy   2147 ATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAATGATAC 2206
Db  37411 ATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAATGATAC 37470

Qy   2207 CTATAAATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTA 2266
Db  37471 CTATAAATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTA 37530

Qy   2267 AGATTTTTTTCTGAAGCCCTAATATTTAAAATGGCCTTATTTTACCATATGGATATAAGA 2326
Db  37531 AGATTTTTTTCTGAAGCCCTAATATTTAAAATGGTCTTATTTTACCATATGGATATAAGA 37590

Qy   2327 TTTGGCTCATAATGATGAGCCCTATCATTTGATTTGAGTTCTATCATTTAAGAGAGCCTA 2386
Db  37591 TTTGGCTCATAATGATGAGCCCTATCATTTGATTTGAGTTCTATCATTTAAGAGAGCCTA 37650

Qy   2387 AATAAAATTATCATCAAGGTATTAAATATAAGACGTTAAATATAATAAAGTGGGGATATA 2446
Db  37651 AATAAAATTATCATCAAGGTATTAAATATAAGACGTTAAATATAATAAAGTGGGGATATA 37710

Qy   2447 TAGAAAACACACAGTGTTAGCACAGAGTAAGATCTCAATGCACATTTGTTGGATGAATAA 2506
Db  37711 TAGAAAACACACAGTGTTAGCACACAGTAAGATCTCAATGCACATTTGTTGGATGAATAA 37770

Qy   2507 ATAAATGCAATTGAATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAGTTCTACTTTA 2566
Db  37771 ATAAATGCAATTGAATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAGTTCTACTTTA 37830

Qy   2567 GAAGTACTAATTGGAGATGACTTTTATATCCCATTTTGGTAATTATTCATACATAGCACA 2626
Db  37831 GAAGTACTAATTGGAGATGACTTTTATATCCCATTTTGGTAATTATTCATACATAGCACA 37890

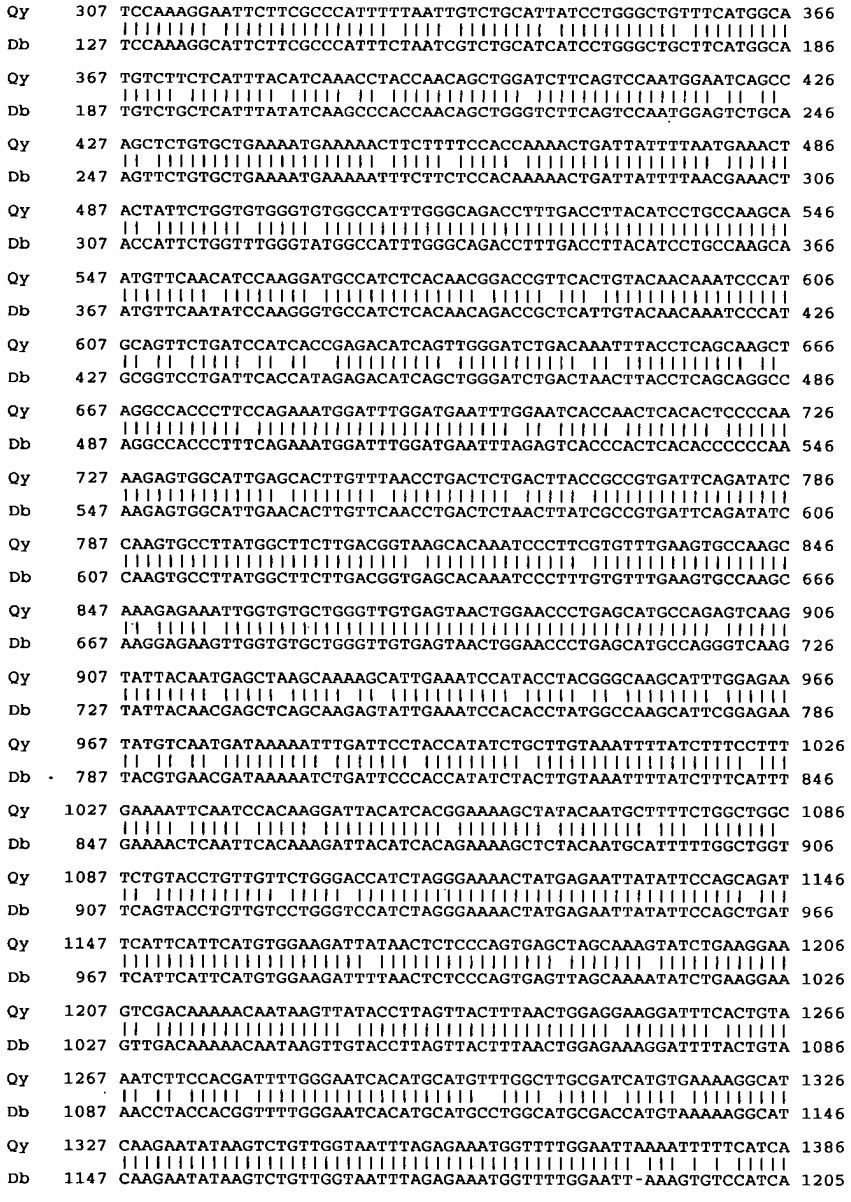
Qy   2627 TATGACCATGATGTTCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 2676
Db  37891 TATGACCATGATGTTCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 37940
```

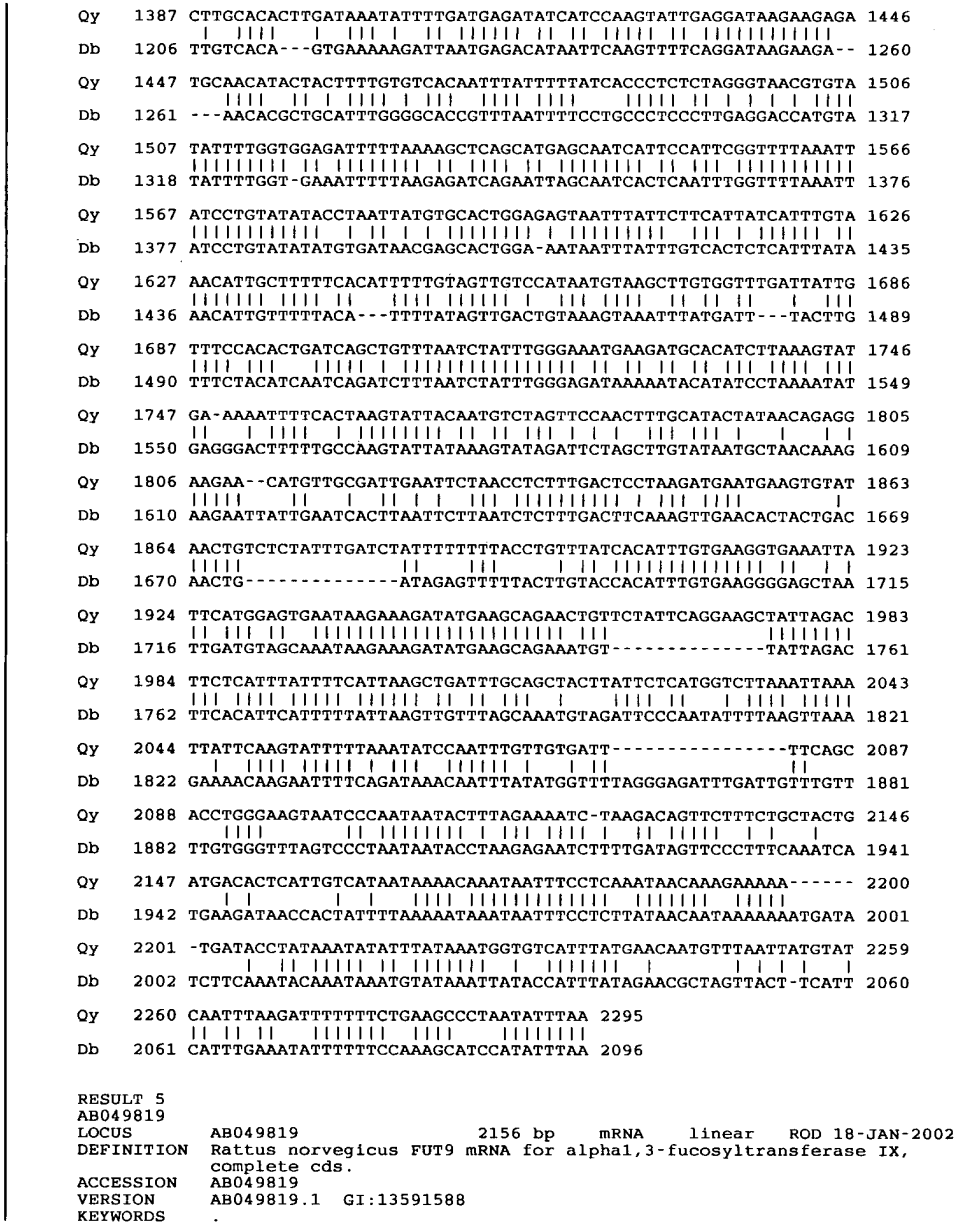
RESULT 4
AB015426
LOCUS AB015426 2139 bp mRNA linear ROD 23-JUN-1999
DEFINITION Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete cds.
ACCESSION AB015426
VERSION AB015426.1 GI:3702718
KEYWORDS Fut9; alpha1,3-fucosyltransferase IX.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K. and Narimatsu,H.
TITLE Expression cloning and characterization of a novel murine alpha1, 3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15) epitope in brain and kidney
JOURNAL J. Biol. Chem. 273 (41), 26729-26738 (1998)
MEDLINE 98434588
REFERENCE 2 (bases 1 to 2139)
AUTHORS Kudo,T. and Narimatsu,H.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science, Soka University, Division of Cell Biology; 1-236 Tangi-cho, Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp, Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES Location/Qualifiers
source 1. .2139
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 115. .1194
/gene="Fut9"
CDS 115. .1194
/gene="Fut9"
/codon_start=1
/product="alpha1,3-fucosyltransferase IX"
/protein_id="BAA33522.1"
/db_xref="GI:3702719"
/translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
CLACDHVKRHQEYKSVGNLEKWFWN"
polyA_site 2139
/note="31 a nucleotides"
BASE COUNT 684 a 395 c 368 g 692 t
ORIGIN

Query Match 45.1%; Score 1206.8; DB 10; Length 2139;
Best Local Similarity 78.4%; Pred. No. 1.1e-233;
Matches 1674; Conservative 0; Mismatches 387; Indels 75; Gaps 16;

```
Qy    187 CCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATA 246
Db      9 CGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGAGTAGTTCA 68

Qy    247 ACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACA 306
Db     69 GCACTCATCTC--TGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACA 126
```





RESULT 5
AB049819
LOCUS AB049819 2156 bp mRNA linear ROD 18-JAN-2002
DEFINITION Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX, complete cds.
ACCESSION AB049819
VERSION AB049819.1 GI:13591588
KEYWORDS .

```
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN
```

```
  Query Match           44.3%;  Score 1184.8;  DB 10;  Length 2156;
  Best Local Similarity 78.2%;  Pred. No. 3.2e-229;
  Matches 1591;  Conservative    0;  Mismatches  392;  Indels   51;  Gaps   12;

Qy      53 CGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGC 112
           | ||  |||| || ||||  ||| | |  || ||    || |   ||||  |||||||||
Db     110 CCCCGGCGCCACTCTCGCCCCCTGGCTACCCACCGATTCCCGCTACGCTGCTAGGACAGC 169

Qy     113 GCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCC 172
           ||||||  | ||||| | ||  | |  || || ||||| | | |||||| || || |
Db     170 GCCGCCGTCACCGCCAGCCCTCGTC--CCGCCCGCGCCTCTCTGCCCTCACGCGCCCTGG 227

Qy     173 GGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATAC 232
               ||  |      || ||| |||||||||||||| |||| ||||| || |||  | ||
Db     228 ATGGAGCTTTCCCTTGCTGGGATCGATTTAGAATGTGATAAGTCAAGTATCTGA--ACAC 285

Qy     233 AGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAA 292
           |||| ||||||  ||||    |||    |||| ||||||||||||||||   |||||||
Db     286 AGTGGAGTAGTTCAACACTCATCT--CAGCTTTCCATGATATGTTCTCTACCGTGAAAAA 343

Qy     293 TTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGG 352
           ||||||| |||||||||||||| |||||||||||||| |||| ||||||||||||||||
Db     344 TTATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGG 403

Qy     353 GCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTC 412
           |||| ||| |||||||||| || || |||||||| || |||||||||||| |||||||||
Db     404 GCTGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTC 463

Qy     413 CAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATT 472
           ||||||| || || |||||||||||||||||||||| ||||| ||||  ||||||||||

Db     464 CAATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATT 523

Qy     473 ATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTA 532
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     524 ATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTA 583

Qy     533 CATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGT 592
           |||||||||||| |||||||||||||||| |||||||||||||||||| ||  |||
Db     584 CATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGT 643

Qy     593 ACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATT 652
           |||||||||||||||| || ||||| || ||| |||||||||| |||||||||| || |
Db     644 ACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACT 703

Qy     653 TACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAA 712
           |||||||||| || ||||| ||||| ||||||||||||||||||||| || || ||||| |
Db     704 TACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCA 763

Qy     713 CTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCC 772
           ||||||| |||||||||||||||||||||||||||| |||||||||| ||||| ||||
Db     764 CTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTAACTTATCGCC 823

Qy     773 GTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGT 832
           |||||||||||| |||||||||| |||||||||||||| ||||||  ||||| ||||
Db     824 GTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGT 883

Qy     833 TTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC 892
           |||||||||||||||||||| || ||||||||||||||||||||||||| || ||||
Db     884 TTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGC 943

Qy     893 ATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGC 952
           ||||||| |||||||||||||| ||||| ||||| || |||||||||| ||||| ||||
Db     944 ATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGC 1003

Qy     953 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 1012
           |||||||||||||||||| |||||||||| ||||||| |||||||| |||||||||
Db    1004 AAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAAT 1063

Qy    1013 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 1072
           |||||||||||||||||| ||||| |||||||||| |||||||| |||||||
Db    1064 TTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATG 1123

Qy    1073 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 1132
           | ||| ||||||| || |||||||||| ||||||||||||||||||||||||||||||||
Db    1124 CATTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAACTATGAGAATT 1183

Qy    1133 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 1192
           |||||||||| |||||||||||||||||||||||||| |||||| |||||||| | ||||
Db    1184 ATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGTGAGTTGGCAA 1243

Qy    1193 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 1252
           | |||||||||| || |||||||||||||||||| |||||||||| ||||||||||
Db    1244 AATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGA 1303

Qy    1253 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 1312
           ||||||| |||||||| || ||||| |||||||||||||||||  |||| |||||||
Db    1304 AGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATC 1363

Qy    1313 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1372
           |||| |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1364 ATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATT 1423

Qy    1373 AAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTG 1432
           ||||| |     ||    |     | |||||  || |||||  | ||| ||||| ||
Db    1424 AAAATGT----CCATCATCGTCACAGTGAAATAGATTAATGAGACACCATTCAAGTGTTC 1479

Qy    1433 AGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTC 1492
           |||||||||||      |||  || | |||| ||||| |||||||||     | ||| |
Db    1480 AGGATAAGAAGA-----AACACGCTGCATTTGGGTCACAGTTTATTTTTCCTGCTCTCCC 1534

Qy    1493 TAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCA 1552
           | | | | | |||||||||||| || |  |||| | |  |           ||||| || |
Db    1535 TTGAGGACCATGTATATTTTGATGAAATTTTTAAGAGACCAGAATCAGCAAATCACTCAA 1594
```

```
Qy    1553 TTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTT 1612
           || ||||||||||||||||||||||  | || |   |||||||| | ||||||||||| |
Db    1595 TTTGGTTTTAAATTATCCTGTATATATGTGATAACAAGCACTGGA-AATAATTTATTCAT 1653

Qy    1613 CATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTG 1672
           |||| |||||||| ||   | || |||     |||| |||||| |  ||| ||||| |||
Db    1654 CATTCTCATTTGTGAAGCATTCTGTTTTTACATTTTATAGTTGACTGTAAAGTAAGTTTG 1713

Qy    1673 TGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGC 1732
           || ||   | ||||||||||||   ||||| | ||| ||||| ||||||||||| || |
Db    1714 TGATT---TAATTGTTTCCACATCAATCAGATCTTTCATCTACTTGGGAAATGAAAATAC 1770

Qy    1733 ACATCTTAAAGTATGAA-AAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCA 1791
           | ||  |||| ||||||  | | ||  | |||| |||  |  | ||| | | |  ||| |
Db    1771 ATATTCTAAAATATGAAGGACTCTTTGCCAAGTGTTATGAAATTTAGATTCTAACTTGTA 1830

Qy    1792 TACTATAACAGAGGAAGAA--CATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGA 1849
           || |   |   |  ||| |   ||   | || | |  ||| |||| ||||| |  ||
Db    1831 TAATGCTAACAAAAAAGGATTATTGAACCACTTAATTCTTAATCTCTCTGACTTCAGAGC 1890

Qy    1850 TGAATGAAGTGTATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTG 1909
           ||||        | || ||                 || |   ||      | || ||  ||||
Db    1891 TGAACACTACTCACAATTG--------------ATGTAGGTTTTACCTGTACCATGTTTG 1936

Qy    1910 TGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCA 1969
           ||| |||| | | ||| ||| |||||| |||||||  | |||||||| |||
Db    1937 TGATGGTGGACTAATTGATGTAGTGAAGAAGAAAGGCAGGAAGCAGAAATGT-------- 1988

Qy    1970 GGAAGCTATTAGACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCAT 2029
             || |||||| | || | |||||||||||| || ||| ||  | |  | |||||
Db    1989 ------TAGTAGACTCCACACTCATTTTCATTAAGTTGTTTTACAAATGCAGA-TCTCAA 2041

Qy    2030 GGTCTTAAATTAAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTT 2083
           | |||| |||||  |   ||| | |   | |||    || |  | |  | |
Db    2042 TATTTTAAGTTAAAGAAAGAAAGAAATACAATTTATGTGGTTGGGGGAGGCTGT 2095
```

```
RESULT 6
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN
```

```
  Query Match           42.9%;  Score 1147.6;  DB 10;  Length 1705;
  Best Local Similarity   83.4%;  Pred. No. 1e-221;
  Matches 1442;  Conservative    0;  Mismatches  259;  Indels   29;  Gaps   11;

Qy     231 ACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAA 290
           |||||||||| ||   ||| | ||||   |||| ||||||   |||||||| | | |||
Db       2 ACAGTGAAGTGGTTCTACACCCGTCT--CTGCTGCCCATG--CTGTTCTCTACAGTAAAA 57

Qy     291 AATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCT 350
           |||||||||||||||||||||||| |||||||| |||||  |||| |||||  |||||||
Db      58 AATTATGACATCAACATCCAAAGGCATTCTTCGTCCATTCCTAATCGTCTGTGTTATCCT 117

Qy     351 GGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAG 410
           || ||| |||||||||||||||| || |||||||||| || ||||||||||| | |||||
Db     118 GGCCTGCTTCATGGCATGTCTGCTTATTTACATCAAGCCCACCAACAGCTGGGTGTTCAG 177

Qy     411 TCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGA 470
           ||||||||| || ||||| |||||||||||||||||||| |||| ||||| ||||||||
Db     178 TCCAATGGAGTCTGCCAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGA 237

Qy     471 TTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCT 530
           |||||||||||||| |||||||| || ||||||||||||||||||||||||||||||||
Db     238 TTATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCT 297

Qy     531 TACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACT 590
           |||||||||||||||| |||||||||||||||| |||||||||| || ||||| |||||
Db     298 TACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACAGACCGCTCACT 357

Qy     591 GTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAA 650
           |||||||||||||||||| || ||||| || ||| |||||||||| |||||||||| ||
Db     358 GTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAA 417

Qy     651 TTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACC 710
           |||||||||| |||||||| ||||| ||||||||||||||||||   | || || ||
Db     418 CTTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAACCTAGAGTCTCC 477

Qy     711 AACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCG 770
           |||||  || |||||||||||||||||||||||||||| ||||| |||||||||| || ||
Db     478 CACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACTCTAACATATCG 537

Qy     771 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGT 830
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db     538 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGT 597

Qy     831 GTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGA 890
           |||||||||||||||||||||||||  |||||||||||||||||||||||||||||| ||
Db     598 GTTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAACTGGAATCCGGA 657

Qy     891 GCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGG 950
           |||||||| |||||||||| |||||||| ||||| || |||||||||||||||||| ||
Db     658 GCATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATCCATACCTATGG 717

Qy     951 GCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAA 1010
           |||||| |||||||||| || ||||| |||||| | ||||| |||||||| |||||||
Db     718 GCAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATATCTACTTGTAA 777

Qy    1011 ATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAA 1070
           |||||||||||| |||||||| ||||| ||||| |||||||||| |||||||| |||||
```

```
Db    778 ATTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAAAAGCTCTACAA 837

Qy   1071 TGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAA 1130
          ||| ||| |||||||||| |||||||||||||||||||||||| |||||||||||||||
Db    838 TGCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAAAACTATGAGAA 897

Qy   1131 TTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
          |||||||||||| ||||||||||||||||||||||||||||||||||||||| | ||
Db    898 TTATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGTTGGC 957

Qy   1191 AAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAG 1250
          ||| ||||| |||||||| |||||||||||||||| |||||||||| || ||||||||
Db    958 AAAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTATTTCAACTGGAG 1017

Qy   1251 GAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGA 1310
          |||||||||||||||||| || ||||||||||||||||||||||||  ||||||| ||
Db   1018 GAAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGCCTGGCTTGTGA 1077

Qy   1311 TCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAA 1370
          ||| || |||||||| ||||| |||||||||| ||||||||||||||||||||||||||
Db   1078 TCACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAATGGTTTTGGAA 1137

Qy   1371 TTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTAT 1430
          |||| ||| |     || |   |    |||| | || |   || |||||  ||||
Db   1138 TTAATGTTTCCCTCATTTTCAAATTTAAGGAAATGTATTAACATGACATCATATAAGTTG 1197

Qy   1431 TGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTC 1490
          | ||||||||||||||   ||||   | ||||    ||||| |||||||| ||  |||||
Db   1198 TAAGGATAAGAAGAGA---AACACCTTGCTTTCAAATCACAGTTTATTTTAATTGCCCTC 1254

Qy   1491 TCTAGGGTAACGTGTATATTTTGGTGGAGA-TTTTTAAAAGCTCAGCATGAGCAATCATT 1549
           || | ||||| |||||||||| ||||| | ||||||| |||||| || |||||||| |
Db   1255 CCTTGAGTAACATGTATATTTTGATGGAAAATTTTTAAGAGCTCAGAATTAGCAATCACT 1314

Qy   1550 CCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATT 1609
          |||| ||  |||||||||||||||||||| | || ||| |||||| | | |||||||||
Db   1315 TCATTTGG-ATTAAATTATCCTGTATATACGTGATAATGAGCACTGAAAA-TAATTTATT 1372

Qy   1610 CTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGC 1669
          || || | ||||||||||||||  |||||||| | || |||||||  ||| |||||
Db   1373 CTCCACTCTCATTTGTAAACATTCTTTTTTCACA-TCTTATAGTTGTCTGTAAAGTAAGT 1431

Qy   1670 TTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGA 1729
          || || ||    |||||||| |||   ||||| | |||| ||| |||||||||| | |
Db   1432 TTATGATT---CAATTGTTTCTACATCAATCAGATCTTTAGTCTGTTTGGGAAATGGAAA 1488

Qy   1730 TGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTG 1789
          ||    || |||| |||||||  ||||  | |||||||| |||||||| | | | ||||
Db   1489 TGGGTGTCCTAAAATATGAAAGCTTTTTGCAAAGTATTATAATGTCTAGATTCTAGTTTG 1548

Qy   1790 CATACTATAACAGAGGAAGAA--CATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAA 1847
          |||| |   |  | |||| |     ||  |  || | |  |||||||||| | |  |
Db   1549 CATAATGGTAACAAAGAAGTATTAGAGTATCACTTAATTCTTAACCTCTTTGAGTTCAGA 1608

Qy   1848 GATGAATGAAGTGTATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATT 1907
          | |||           ||||| | || ||||| ||   |||||| ||||||| ||||||
Db   1609 GTTGA----------CACTGTGTGTAATTGATGTA--CTTTTTAACTGTTTACCACATT 1655

Qy   1908 TGTGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAA 1957
          |||||||    | | | |||||  || ||||||||||||||||||||||
Db   1656 TGTGAAGACAGACTGACTCATGTTGTAAATAAGAAAGATATGAAGCAGAA 1705


RESULT 7
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN

  Query Match           35.8%;  Score 958.2;  DB 10;  Length 1128;
  Best Local Similarity   91.2%;  Pred. No. 1.7e-183;
  Matches 1017;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy    260 TGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTC 319
          ||||| |||||||||||||||   |||||||||||||| ||||||||||||||| ||||
Db     14 TGCTTTCCATGATATGTTCTCTACCGTGAAAAATTATGACCTCAACATCCAAAGGCATTC 73

Qy    320 TTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTT 379
          |||||||||| |||| ||||||||||||||||||| ||| |||||||||| || || |
Db     74 TTCGCCCATTTCTAATCGTCTGCATTATCCTGGGCTGCTTCGTGGCATGTCTACTTATCT 133

Qy    380 ACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGA 439
          ||||||| || |||||||||||| |||||||||||||||| || || ||||||||||||
Db    134 ACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCAAGCTCTGTGCTGA 193

Qy    440 AAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGT 499
          |||||||||| ||||| ||||  |||||||||||||||||||||||||||||||||||| |
Db    194 AAATGAAAAATTTCTTCTCCAGGAAAACTGATTATTTTAATGAAACTACTATTCTGGTTT 253

Qy    500 GGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCC 559
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    254 GGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCCATGTTCAACATCC 313

Qy    560 AAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCC 619
          |||| |||||||||||||||||| ||  |||||||||||||||||| || ||||| |
Db    314 AAGGGTGCCATCTCACAACGGACCGCTCGTTGTACAACAAATCCCATGCGGTCCTGATAC 373

Qy    620 ATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCC 679
          | ||| |||||||||| |||||||||| || |||||||||| || ||||| ||||| |
Db    374 ACCACAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTC 433

Qy    680 AGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTG 739
          |||||||||||||||||| || || ||||| |||||||| ||||||||||||||||||||
Db    434 AGAAATGGATTTGGATGAACTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTG 493
```

```
Qy    740 AGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATG 799
          |||||||||| |||||||||| ||||| |||||||||||||||| |||||||||| |
Db    494 AGCACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATACAAGTGCCTTACG 553

Qy    800 GCTTCTTGACGGTAAGCACAAATCCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGG 859
          |||||||||||| ||||||| ||||| |||||||||||||||||||||||| || |
Db    554 GCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAG 613

Qy    860 TGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGC 919
          |||||||||||||||||||||||| || |||||||||| |||||||||||||| ||||
Db    614 TGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGC 673

Qy    920 TAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATA 979
          | ||||| || |||||||||| ||||| |||||||||||||||||||||| |||||||
Db    674 TCAGCAAGAGTATTGAAATCCACACCTATGGGCAAGCATTTGGAGAATATGTGAATGATA 733

Qy    980 AAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCC 1039
          ||||| ||||||| |||||||| |||||||||||||||||||||||||||| ||||| |
Db    734 AAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTC 793

Qy   1040 ACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1099
          |||| |||||||||| |||||||| |||||||| ||| ||||||| ||||||||||||
Db    794 ACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGTTCTGTACCTGTTG 853

Qy   1100 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1159
          | ||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    854 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 913

Qy   1160 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACA 1219
          ||||||||| |||||||||||||||| | ||||| |||||||||| || ||||||||||
Db    914 TGGAAGATTTTAACTCTCCCAGTGAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACA 973

Qy   1220 ATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGAT 1279
          |||||||||| |||||||||| |||||||||||||||| |||||||| || ||||| |
Db    974 ATAAGTTATATCTTAGTTACTTCAACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGT 1033

Qy   1280 TTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGT 1339
          |||||||||||||||||  |||| |||||||||| ||||||||||||||||||||||||
Db   1034 TTTGGGAATCACATGCATGCCTGGCATGCGATCATGTAAAAAGGCATCAAGAATATAAGT 1093

Qy   1340 CTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1374
          ||||||| ||||||||||||||||||||||||||
Db   1094 CTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1128


RESULT 8
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN
```

Query Match 30.7%; Score 820.4; DB 5; Length 1707;
Best Local Similarity 76.7%; Pred. No. 1.3e-155;
Matches 1061; Conservative 1; Mismatches 302; Indels 20; Gaps 4;

```
Qy    288 AAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTAT 347
          | |||||||||||||||||| ||||||||| | || |||||||| ||||||| |||| |
Db    261 ACAAATTATGACATCAACATCTAAAGGAATTTTCCGGCCATTTTTTATTGTCTTCATTGT 320

Qy    348 CCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTT 407
          :|| || |||||||||||||| | | || ||||| || ||||| || | ||||||||||
Db    321 YCTTGGTTGTTTCATGGCATTTATACTAATTTATATTAAACCAACAAGCAGCTGGATCTC 380

Qy    408 CAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAAC 467
          | |||  |||||||||||||||| ||  |||||||||| | |||||||||  |||||||
Db    381 TAATCCTGTGGAATCAGCCAGCTCAGTTTTGAAAATGAAGAGCTTCTTTTCTTCCAAAAC 440

Qy    468 TGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGA 527
          |||| ||||||||||||||| |||| |||| ||| | |||||||||||| ||||| || ||
Db    441 TGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGACATTCGA 500

Qy    528 CCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTC 587
          || |||||||||||| | |||||||||||||| || ||||||| || |  |||| ||
Db    501 TCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTGACCGCTC 560

Qy    588 ACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGAC 647
          ||| || |||| |||||||||||||| || |||||| | ||||| |  ||||||||||
Db    561 ACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGGATCTGAC 620

Qy    648 AAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATC 707
          |||||||||||||||| |||||||| |||||||||||||||||||||| ||||||||
Db    621 TAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACTTGGAATC 680

Qy    708 ACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTA 767
          ||||||||  ||||| |||||||||||||||| ||| | |||||||| || ||||||||
Db    681 TCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCCTGACTTA 740

Qy    768 CCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTT 827
          ||| |||||||||||||| || |||||||||||| | |||  ||  | |||    ||||
Db    741 CCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAGGCGCCTT 800

Qy    828 CGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCC 887
            | |||||||||||| ||||| || ||||| || |||||||| ||||||||||||||
Db    801 TACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACTGGAACCC 860

Qy    888 TGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTA 947
          |||||||||  |||||||||||||||||||||||| |||||   ||||||||||||||||
Db    861 TGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCCATACCTA 920

Qy    948 CGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTG 1007
          ||||||||| |||||||| || ||||| |||||||| || ||||| || || ||   |||
Db    921 CGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCTCCAGTTG 980

Qy   1008 TAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATA 1067
```

```
Db    981 CAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGAAACTTTA 1040

Qy   1068 CAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGA 1127

Db   1041 CAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAAACTACGA 1100

Qy   1128 GAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCT 1187

Db   1101 GAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAAGAGAGCT 1160

Qy   1188 AGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 1247

Db   1161 TGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACTTCAACTG 1220

Qy   1248 GAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTG 1307

Db   1221 GAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTCTTGCTTG 1280

Qy   1308 CGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTG 1367

Db   1281 TGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAATGGTTTTG 1340

Qy   1368 GAATTAA--AATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCA 1425

Db   1341 GAACTAATTTGTGTGTGTGTATGTGCACTTTTTTGAAGAAGATAATGGATCTTTAGTCTA 1400

Qy   1426 AGTATTGAGGATAAGAA----------------GAGATGCAACATACTACTTTTGTGTCA 1469

Db   1401 AGTGGAGAGTAAAGGAATAAGGAAAAAAGAAGGAAGGATGAACCTCACAGAACCTTACGT 1460

Qy   1470 CAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAA 1529

Db   1461 CAAGGTTTATCAGCCATCCTCTCTTGGCTATTCTATTTATATTTTGTAAGAGATTTTAAA 1520

Qy   1530 GCTCAGCATGAGCAATCAT-TCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATG 1588

Db   1521 AATTATGAGCAGCAGTCATCAAAACTCAGTTTTAAATTATAATGTACATA-CTAATTATG 1579

Qy   1589 TGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTT 1648

Db   1580 TGCACTGAAGAATATATGATTGTTTACTATAATTTTTAAACAAGATTTTTCACATTTTCT 1639

Qy   1649 GTAG 1652

Db   1640 GTGG 1643
```

```
RESULT 9
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN
```

```
  Query Match             25.3%;  Score 676.8;  DB 5;  Length 1080;
  Best Local Similarity   76.7%;  Pred. No. 1.2e-126;
  Matches  828;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy    295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354

Db      1 ATGACATCAGCATCTAAAGGAATCTTACGGGCATTTTTAGCTGTCAGCGGTGTGTTGATC 60

Qy    355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414

Db     61 TGTTTTATGATATGTTTACTGATATACGTCAAACCAACAAACAACTGGATTTCCAGTCCT 120

Qy    415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474

Db    121 ATCGAATCAGCACATTCAGTCCTTAAAATGAAGAATTTTTTCTCCTCTAAACAGAATTAT 180

Qy    475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534

Db    181 TACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGACTTTTGAACTGAAA 240

Qy    535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594

Db    241 TCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTGACCGCGCACTCTAC 300

Qy    595 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 654

Db    301 AACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGGATCTCATTAACCTG 360

Qy    655 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 714

Db    361 CCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATCTGGAATCACCAACA 420

Qy    715 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 774

Db    421 CATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTCTGACTTACAGGCGT 480

Qy    775 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834

Db    481 GACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAAAACCATTTGACTTT 540

Qy    835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894

Db    541 GAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACTGGAACCCAGAACAT 600

Qy    895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954

Db    601 GCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTATAACTTATGGGCAA 660

Qy    955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014

Db    661 GCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTGCATCTTGCAAATTT 720
```

```
Qy    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 1074
           ||||| || ||||||||||| || ||||| ||||| ||||| ||||| || || |||||
Db     721 TATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAAAACTTTATAATGCA 780

Qy    1075 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1134
           ||  ||||||| || || ||  |||| |||||||  |||| |||||||||||||| ||
Db     781 TTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAAACTATGAGAACTAC 840

Qy    1135 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1194
           ||||| |||||||| || || || |||||||||| |  |||||| || ||  ||||  |
Db     841 ATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAAGAGAATTAGCTGAC 900

Qy    1195 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
            ||||        |  ||||| | |    |  ||||| || ||||  ||||||||||
Db     901 CATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACTACAACTGGAGGAAA 960

Qy    1255 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1314
            |||| || || ||| | || |  || |||||||| ||||||| || || || || |||
Db     961 CATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCTTAGCGTGTGACCAT 1020

Qy    1315 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1374
           ||||||||||||||||||| || || || || ||| ||||||| ||||| ||||| |||
Db    1021 GTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGTGGTTCTGGAACTAA 1080
```

RESULT 10
AC100674
LOCUS AC100674 71396 bp DNA linear HTG 22-NOV-2001
DEFINITION Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC100674
VERSION AC100674.1 GI:17048040
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 71396)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-167K24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 71396)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L15932
Center clone name: 167_K_24

* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 788: contig of 788 bp in length
* 789 888: gap of 100 bp
* 889 1702: contig of 814 bp in length
* 1703 1802: gap of 100 bp
* 1803 2642: contig of 840 bp in length
* 2643 2742: gap of 100 bp
* 2743 3637: contig of 895 bp in length
* 3638 3737: gap of 100 bp
* 3738 4635: contig of 898 bp in length
* 4636 4735: gap of 100 bp
* 4736 5657: contig of 922 bp in length
* 5658 5757: gap of 100 bp
* 5758 6692: contig of 935 bp in length
* 6693 6792: gap of 100 bp
* 6793 7703: contig of 911 bp in length
* 7704 7803: gap of 100 bp
* 7804 8742: contig of 939 bp in length
* 8743 8842: gap of 100 bp
* 8843 9787: contig of 945 bp in length
* 9788 9887: gap of 100 bp
* 9888 10819: contig of 932 bp in length
* 10820 10919: gap of 100 bp
* 10920 11808: contig of 889 bp in length
* 11809 11908: gap of 100 bp
* 11909 12830: contig of 922 bp in length
* 12831 12930: gap of 100 bp
* 12931 13847: contig of 917 bp in length
* 13848 13947: gap of 100 bp
* 13948 14886: contig of 939 bp in length
* 14887 14986: gap of 100 bp
* 14987 15911: contig of 925 bp in length
* 15912 16011: gap of 100 bp
* 16012 16903: contig of 892 bp in length
* 16904 17003: gap of 100 bp
* 17004 17906: contig of 903 bp in length
* 17907 18006: gap of 100 bp
* 18007 18931: contig of 925 bp in length
* 18932 19031: gap of 100 bp
* 19032 19945: contig of 914 bp in length
* 19946 20045: gap of 100 bp
* 20046 20951: contig of 906 bp in length
* 20952 21051: gap of 100 bp
* 21052 21949: contig of 898 bp in length
* 21950 22049: gap of 100 bp
* 22050 22950: contig of 901 bp in length
* 22951 23050: gap of 100 bp
* 23051 23945: contig of 895 bp in length
* 23946 24045: gap of 100 bp
* 24046 24955: contig of 910 bp in length
* 24956 25055: gap of 100 bp
* 25056 25980: contig of 925 bp in length
* 25981 26080: gap of 100 bp
* 26081 27052: contig of 972 bp in length
* 27053 27152: gap of 100 bp
* 27153 28067: contig of 915 bp in length
* 28068 28167: gap of 100 bp
* 28168 29076: contig of 909 bp in length
```
```

```
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp

  Query Match             23.9%;  Score 639;  DB 2;  Length 71396;
  Best Local Similarity   85.0%;  Pred. No. 1.3e-118;
  Matches  764;  Conservative    0;  Mismatches  122;  Indels   13;  Gaps    4;

Qy    366 ATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGC 425
          || |||   | | |     |  || |||||||||| | ||||||||||||||||| || ||
Db  65456 ATCTCTCGCCCTGTGGTGGATTCCCACCAACAGCTCGGTCTTCAGTCCAATGGAGTCTGC 65515

Qy    426 CAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAAC 485
          || ||||||||||||||||||| ||||| ||||| |||||||||||||||||| |||||
Db  65516 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 65575

Qy    486 TACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 545
          ||| |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db  65576 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 65635

Qy    546 AATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCA 605
          ||||||||| |||||||| |||||||||||||| |||| ||| ||||||||||||||||
Db  65636 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 65695

Qy    606 TGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGC 665
          ||| || ||||| || ||  |||||||||| |||||||||| || ||||||||||| ||
Db  65696 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 65755

Qy    666 TAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCA 725
          |||||||||| |||||||||||||||||||||| || ||||| |||||||| |||||
Db  65756 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 65815

Qy    726 AAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATAT 785
          |||||||||||||| |||||||| |||||||||| ||||| ||||||||||||||||||
Db  65816 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 65875

Qy    786 CCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAG 845
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db  65876 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 65935

Qy    846 CAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAA 905
          ||| ||||| ||||||||||||||||||||||||||||||||||||||||||| |||||
Db  65936 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 65995

Qy    906 GTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGA 965
          ||||||||| ||||| ||||| || |||||||||| ||||| || |||||||| |||||
Db  65996 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 66055

Qy    966 ATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTT 1025
          ||| || || |||||||| ||||||| |||||||| |||||||||||||||||||  |
Db  66056 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCTCT 66115

Qy   1026 TGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGG 1085
            ||| |      ||  | | | |  | |         | ||     ||| |||||||
Db  66116 GAAAACTCATTCACAAGATTACTCCCAAAAG---------TCCACATGCTTTTTGGCTGG 66166

Qy   1086 CTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGA 1145
          || |||||||||| ||||| ||||||||||||||||||||||||||||||||||||| ||
Db  66167 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 66226

Qy   1146 TTCATTCATTC--ATGTGGAAGATTATAACTCT-CCCAGTGAGCTAGCAAAGTATCTGAA 1202
          |||||||||||    |||||||||| | ||||| |||||||| |||||||    |||||
Db  66227 TTCATTCATTCCATCGTGGAAGATTTTCACTCTCCCCAGTGAGTTAGCAAATATCCTGAA 66286

Qy   1203 GGAAGTCGACAAAAACAATAAGTT-ATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 1260
          |||||||||||| ||||||||||  |||||||||||| ||||| |  ||| |  |||
Db  66287 GGAAGTCGACAANAACAATAAGTTGGTACCTTAGTTACTCTAACTCGNAGAAAGGATTC 66345
```

```
RESULT 11
AC100674/c
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1702: contig of 814 bp in length
            *     1703 1802: gap of      100 bp
            *     1803     2642: contig of 840 bp in length
            *     2643 2742: gap of      100 bp
            *     2743     3637: contig of 895 bp in length
            *     3638 3737: gap of      100 bp
            *     3738     4635: contig of 898 bp in length
            *     4636 4735: gap of      100 bp
            *     4736     5657: contig of 922 bp in length
            *     5658 5757: gap of      100 bp
            *     5758     6692: contig of 935 bp in length
            *     6693 6792: gap of      100 bp
            *     6793     7703: contig of 911 bp in length
            *     7704 7803: gap of      100 bp
            *     7804     8742: contig of 939 bp in length
            *     8743 8842: gap of      100 bp
            *     8843     9787: contig of 945 bp in length
            *     9788 9887: gap of      100 bp
            *     9888    10819: contig of 932 bp in length
            *    10820 10919: gap of      100 bp
            *    10920    11808: contig of 889 bp in length
            *    11809 11908: gap of      100 bp
            *    11909    12830: contig of 922 bp in length
            *    12831 12930: gap of      100 bp
            *    12931    13847: contig of 917 bp in length
            *    13848 13947: gap of      100 bp
            *    13948    14886: contig of 939 bp in length
            *    14887 14986: gap of      100 bp
            *    14987    15911: contig of 925 bp in length
            *    15912 16011: gap of      100 bp
            *    16012    16903: contig of 892 bp in length
            *    16904 17003: gap of      100 bp
            *    17004    17906: contig of 903 bp in length
            *    17907 18006: gap of      100 bp
            *    18007    18931: contig of 925 bp in length
            *    18932 19031: gap of      100 bp
            *    19032    19945: contig of 914 bp in length
            *    19946 20045: gap of      100 bp
            *    20046    20951: contig of 906 bp in length
            *    20952 21051: gap of      100 bp
            *    21052    21949: contig of 898 bp in length
            *    21950 22049: gap of      100 bp
            *    22050    22950: contig of 901 bp in length
            *    22951 23050: gap of      100 bp
            *    23051    23945: contig of 895 bp in length
            *    23946 24045: gap of      100 bp
            *    24046    24955: contig of 910 bp in length
            *    24956 25055: gap of      100 bp
            *    25056    25980: contig of 925 bp in length
            *    25981 26080: gap of      100 bp
            *    26081    27052: contig of 972 bp in length
            *    27053 27152: gap of      100 bp
            *    27153    28067: contig of 915 bp in length
            *    28068 28167: gap of      100 bp
            *    28168    29076: contig of 909 bp in length
            *    29077 29176: gap of      100 bp
            *    29177    30093: contig of 917 bp in length
            *    30094 30193: gap of      100 bp
            *    30194    31099: contig of 906 bp in length
            *    31100 31199: gap of      100 bp
            *    31200    32097: contig of 898 bp in length
            *    32098 32197: gap of      100 bp
            *    32198    33113: contig of 916 bp in length
            *    33114 33213: gap of      100 bp
            *    33214    34125: contig of 912 bp in length
            *    34126 34225: gap of      100 bp
            *    34226    35133: contig of 908 bp in length
            *    35134 35233: gap of      100 bp
            *    35234    36150: contig of 917 bp in length
            *    36151 36250: gap of      100 bp
            *    36251    37171: contig of 921 bp in length
            *    37172 37271: gap of      100 bp
            *    37272    38159: contig of 888 bp in length
            *    38160 38259: gap of      100 bp
            *    38260    39156: contig of 897 bp in length
            *    39157 39256: gap of      100 bp
            *    39257    40188: contig of 932 bp in length
            *    40189 40288: gap of      100 bp
            *    40289    41210: contig of 922 bp in length
```

```
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp
```

Query Match 12.8%; Score 343.4; DB 2; Length 71396;
Best Local Similarity 64.3%; Pred. No. 7.2e-59;
Matches 633; Conservative 0; Mismatches 303; Indels 49; Gaps 12;

```
Qy   1102 CTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTG 1161
          | | | | |   |  |
Db  30210 CAGTGGCGAGACATCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 30151

Qy   1162 GAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAAT 1221
                                                                  |||
Db  30150 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAT 30091

Qy   1222 AAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCAC-GATT 1280
          ||||  |||||||||||||||||||||||| |||||||| |||||||| || |||| | ||
Db  30090 AAGTNGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGGTT 30031

Qy   1281 TTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTC 1340
          |||||||||||||||||  |||| ||||| ||||| ||||||||||||||||||||||||
Db  30030 TTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTC 29971

Qy   1341 TGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGAT 1400
          |||||||||||||||||||||||||||||||||||| ||| | | ||||| |  |||   |
Db  29970 TGTTGGTAATTTAGAGAAATGGTTTTGGAATT-AAAGTGTCCATCATTGTCACA---GTG 29915

Qy   1401 AAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACT 1460
          ||| |  || |||||| || || ||||| || ||||||||||||     ||||  || |
Db  29914 AAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA-----AACACGCTGCA 29860

Qy   1461 TTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGA 1520
          |||| | |||  |||| ||||    ||||| || | | | | |||||||||||||| || |
Db  29859 TTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGT-GAAA 29801

Qy   1521 TTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACC 1580
          ||||||| || |||| || |||||||| || ||| ||||||||||||||||||||||||
Db  29800 TTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATG 29741

Qy   1581 TAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTC 1640
          | || | | |||||||| | |||||||||  ||| | |||||| |||||||| |||| |
Db  29740 TGATAACGAGCACTGGA-AATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTAC 29682

Qy   1641 ACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTGTTTCCACACTGATC 1700
          |   |||| |||||| |  ||| ||||  || || |||     ||||||| |||   |||
Db  29681 A---TTTTATAGTTGACTGTAAAGTAAATTTATGATTTAC---TTGTTTCTACATCAATC 29628

Qy   1701 AGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTATGA-AAAATTTTCAC 1759
          || | |||||||||||||||| || || || || ||| |||| |||||   | ||||  |
Db  29627 AGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATATGAGGGACTTTTTGC 29568

Qy   1760 TAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGGAAGAA--CATGTTG 1817
           |||||||| || || ||| | | |  ||| ||| |   |   | ||||||    ||
Db  29567 CAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAGAAGAATTATTGAAT 29508

Qy   1818 CGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTGTATAACTGTCTCTATTT 1877
          |  || | |  ||| ||||||||||| | ||| ||||        | |||||
Db  29507 CACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGACAACTG--------- 29457

Qy   1878 GATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAATTATTCATGGAGTGAAT 1937
             ||    |||      | || ||||||||||||||| ||  | ||| ||| ||  |||
Db  29456 -----ATAGAGTTTTTACTTGTACCACATTTGTGAAGGGGAGCTAATTGATGTAGCAAAT 29402

Qy   1938 AAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTAGACTTCTCATTTATTTT 1997
          |||||||||||||||||||| |||                ||||||||||| |||| |||||
Db  29401 AAGAAAGATATGAAGCAGAAATGT--------------TATTAGACTTCACATTCATTTT 29356

Qy   1998 CATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATTAAATTATTCAAGTATTT 2057
          ||||||  || || |||  |    |||| ||   | |||| |||||  |  |||| ||||
Db  29355 TATTAAGTTGTTTAGCAAATGTAGATTCCCAATATTTTAAGTTAAAGAAAACAAGAATTT 29296

Qy   2058 TTAAATATCCAATTTGTTGTGATTT 2082
          | | |||  |||||| |   | |||
Db  29295 TCAGATAAACAATTTATATGGTTTT 29271
```

```
RESULT 12
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
```

```
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN

  Query Match             10.4%;  Score 279.6;  DB 5;  Length 1567;
  Best Local Similarity   58.3%;  Pred. No. 2.5e-46;
  Matches  533;  Conservative    0;  Mismatches  369;  Indels   12;  Gaps    2;

Qy    477 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 536
          | | ||||| |   |  || | ||||||| || || ||       |||||| |   |
Db    461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy    537 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 596
            ||     |  ||||||||||||| || || ||| | |||   ||| |     ||  | |
Db    521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy    597 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 656
           ||| | |||| |||| || | || ||| |||| | ||  ||       |||| ||
Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGG---------ATTTGCC 631

Qy    657 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 716
           || |   || | ||| | ||||| || ||||| |||  |||| |||| || || | | |
Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy    717 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 776
            | |  | ||| ||    ||  |  |   |||||| |||||    | || ||||   ||
Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy    777 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTT 833
          |||||||||  | |||||||||    |  |  || |     |  | |     | |         ||
Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy    834 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 893
            | ||  |  || || ||| | || ||||||  | ||  |||||||    |     | ||
Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy    894 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 953
            |     |||| || |||||| |    |||  |||| || |  |  |||| ||
Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy    954 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 1013
            ||      |    || |||||| |  |||        |    ||||   ||||||||||
Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy   1014 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 1073
          || |||||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy   1074 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 1133
          |||   |   ||  | || |||||||| || || || || || || |||||||||| | |
Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy   1134 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 1193
          ||| |||  ||| | ||||| |||||||| ||||      ||||||  || || ||  |
Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy   1194 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 1253
            | ||  ||     | ||| ||||| |  |    |||       || || |||||||| ||
Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy   1254 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 1313
            |||| | |  ||    | ||  |||| ||     ||||| |||  | |||| ||| |
Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy   1314 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1373
          | | |||||| || |  | |||    || ||    || | |    |||| ||| |||
Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy   1374 AAATTTTTCATCAC 1387
          |     ||  ||||
Db   1352 ATGGGATTGTTCAC 1365


RESULT 13
AB023628
LOCUS       AB023628                1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
```

```
                  LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                  *
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match             8.8%;  Score 234.6;  DB 5;  Length 1832;
  Best Local Similarity   55.5%;  Pred. No. 3.2e-37;
  Matches  500;  Conservative    0;  Mismatches  389;  Indels   12;  Gaps    2;

Qy    477 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 536
          | | || |  |   |  || | ||| ||   || ||  |     | |||||||| | |
Db    619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy    537 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 596
           ||       ||  |||||||| || || ||  | |||   ||| |    |||  | |
Db    679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy    597 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 656
          ||| | |||| |||| || | || ||| |||| || ||           ||| |||| ||
Db    739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGC---------ACAGATTTTCC 789

Qy    657 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 716
          || |   | | |||   ||||| || ||||| |||  |||| |||| || || || ||
Db    790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy    717 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 776
            |||||      | |  |  |   ||||||  ||||    | ||||  | | ||
Db    850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy    777 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA---AATCCCTTCGTGTT 833
          ||||||||| | |||||||||| ||      |    |    ||||    | |        |||
Db    910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy    834 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 893
          | || |  || || ||| | || ||||||  | || |||||||    |     | ||
Db    970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy    894 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 953
            |     | |  | | | ||| |    |||  |||| |  |    ||||| ||
Db   1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy    954 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 1013
            ||      |     | |||| | |  |||       |  | ||||   ||||||||
Db   1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy   1014 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 1073
          || || |||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db   1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy   1074 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 1133
          |||       ||  | || |||||  ||  | || ||  | ||  | |||||||  | |
Db   1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy   1134 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 1193
           || |||  |||  | ||||| |||||||| ||||      ||||||  || || ||  |
Db   1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy   1194 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 1253
            | ||  |     | ||| ||||| |  |    |||       |  || ||||||||||
Db   1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy   1254 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 1313
            ||    |  |||   | ||   ||| ||     |||| | |||  |  ||| ||| |
Db   1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy   1314 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1373
          | | |||||| || |  | ||||  | |||    || | |    ||||| |||  |||
Db   1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy   1374 A 1374
          |
Db   1510 A 1510
```

```
RESULT 14
AB035907
LOCUS       AB035907               1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN

  Query Match             8.3%;  Score 222.6;  DB 5;  Length 1092;
  Best Local Similarity   57.3%;  Pred. No. 7.6e-35;
  Matches  484;  Conservative    0;  Mismatches  349;  Indels   12;  Gaps    4;

Qy    478 AATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 537
          ||  ||   ||| |  |  | |||     ||||||||  |        |||
Db    199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy    538 TGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAAC 597
          |||||     | |||||  ||   || |||||  |||||||  || |    || ||| |
Db    259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy    598 AAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCT 657
            |  |  | ||  || || |||| ||||| |||||   ||  ||| | |   |  | |||
Db    319 GAGGCAGACGCCATTATGTTCCACCACCGGGACAT---TTCTGATTTCAGTGACCTTCCT 375

Qy    658 CAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 717
            | |    | |    || ||| || ||||| |||||||| || ||||| ||  | ||
Db    376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy    718 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 777
            | |||     |    |  | | |  | | || |||  |||  || | ||| |    ||
Db    436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy    778 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 837
          || ||||||    |||| || ||||||||      |   ||  || | |    |  | |
```

```
Db    496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy    838 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 897

Db    553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy    898 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 957

Db    613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy    958 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 1017

Db    670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy   1018 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 1074

Db    730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy   1075 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1134

Db    790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy   1135 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1194

Db    850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy   1195 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Db    910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy   1255 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1314

Db    970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy   1315 GTGAA 1319

Db   1030 GTCAA 1034
```

```
RESULT 15
AF288369
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN
```

```
  Query Match             8.2%;  Score 220.2;  DB 5;  Length 1264;
  Best Local Similarity   58.0%;  Pred. No. 2.4e-34;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    6;  Gaps    3;

Qy    487 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 546

Db    175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy    547 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606

Db    235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy    607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 666

Db    295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy    667 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 726

Db    355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy    727 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 786

Db    415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy    787 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 846

Db    475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy    847 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 906

Db    535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy    907 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 966

Db    595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGC-GGCAACACTTGCC 653

Qy    967 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 1026

Db    654 CCTGTCCATG--GACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy   1027 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATAC---AATGCTTTTCTGGCT 1083

Db    712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy   1084 GGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCA 1143

Db    772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy   1144 GATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAG 1203

Db    832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy   1204 GAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Db    892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942
```

```
Search completed: October  6, 2002, 04:42:32
Job time : 3757.32 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 300.303 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-5
Perfect score: 2676
Sequence: 1 gcgcgcgcggcgcagcagct.....ccaaataaacctaccattac 2676

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2676	100.0	2676	21	AAZ92647	Human alpha-1,3-fu
2	1542	57.6	2822	21	AAZ92646	Human alpha-1,3-fu
3	1206.8	45.1	2170	21	AAZ92645	Murine alpha-1,3-f
4	1159.2	43.3	2036	21	AAZ92672	Murine alpha-1,3-f
5	127.4	4.8	1814	18	AAT59505	Murine myeloid-lin
6	127	4.7	3594	18	AAT85219	Mouse alpha-fucosy
7	114.4	4.3	1701	15	AAQ77732	Human alpha-1,3-fu
8	114.4	4.3	1701	21	AAF21136	Human low adenosin
9	114.4	4.3	1701	21	AAA35014	Human adenosine re
10	114.4	4.3	6944	21	AAF21137	Human low adenosin
11	114.4	4.3	6944	21	AAA35015	Human adenosine re
12	113.4	4.2	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
13	113.4	4.2	1654	18	AAT61680	Human alpha(1,3)-f
14	112.2	4.2	1086	15	AAQ56912	Lewis blood group
15	112.2	4.2	2042	18	AAT80111	Fucosyltransferase
16	112.2	4.2	2043	12	AAQ13330	Glycosyltransferas
17	112.2	4.2	2043	15	AAQ56906	DNA encoding a gly
18	112.2	4.2	2043	18	AAT76769	Human alpha 1,3/4
19	112.2	4.2	2043	18	AAT61675	Human alpha(1,3/1,
20	111.8	4.2	1126	21	AAF21135	Human low adenosin
21	111.8	4.2	1126	21	AAA35013	Human adenosine re
22	111.8	4.2	1316	15	AAQ56911	DNA encoding a gly
23	111.8	4.2	1316	18	AAT61679	Human alpha(1,3)-f
24	106.2	4.0	347	14	AAQ59752	Human brain Expres
25	103.6	3.9	795	17	AAT13798	Alpha-(1,3/1/4)-fu
26	102.4	3.8	2175	11	AAQ06691	CDX, a MILA involv
27	100.8	3.8	1256	21	AAF21134	Human low adenosin
28	100.8	3.8	1256	21	AAA35012	Human adenosine re
29	100.8	3.8	1400	13	AAQ31436	Encodes a HeLa cel
30	100.8	3.8	1488	15	AAQ56910	DNA encoding a gly
31	100.8	3.8	2175	12	AAQ14382	Clone 7.2 encoding
32	100.8	3.8	2175	24	AAS17082	Human cDNA clone 7
33	100.8	3.8	2861	12	AAQ14383	Clone 1 encoding 1
34	100.8	3.8	2861	21	AAF21133	Human low adenosin
35	100.8	3.8	2861	21	AAA35011	Human adenosine re
36	100.8	3.8	2861	24	AAS17083	Human cDNA clone 1
37	100.8	3.8	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
38	100.8	3.8	3647	15	AAQ56909	DNA encoding a gly
39	100.8	3.8	3647	18	AAT61678	Human alpha(1,3)-f
40	97.6	3.6	2134	18	AAT59506	Human myeloid deri
41	77.4	2.9	90	21	AAC24928	Human secreted pro
c 42	69.6	2.6	883	22	AAL15210	Human breast cance
43	67.4	2.5	5979	22	AAS45313	Chemically pretrea
44	66.6	2.5	1578	23	ABL11867	Drosophila melanog
c 45	66.6	2.5	4229	23	ABL11866	Drosophila melanog

ALIGNMENTS

RESULT 1
AAZ92647
ID AAZ92647 standard; cDNA; 2676 BP.
XX
AC AAZ92647;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 295..1374
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
DR P-PSDB; AAY80996.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Claim 4; Page 143-150; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC AAZ92646 and AAZ92647 represent cDNAs encoding human
CC alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

Query Match 100.0%; Score 2676; DB 21; Length 2676;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCG 60

Qy      61 CCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGCCGCCAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGCCGCCAC 120

Qy     121 CGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGGATGGCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGGATGGCG 180

Qy     181 CTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGT 240

Qy     241 AGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACA 300

Qy     301 TCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTC 360

Qy     361 ATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAA 420

Qy     421 TCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAAT 480

Qy     481 GAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGC 540

Qy     541 CAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAA 600

Qy     601 TCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAG 660

Qy     661 CAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACT 720

Qy     721 CCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCA 780

Qy     781 GATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTG 840

Qy     841 CCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGA 900

Qy     901 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 960

Qy     961 GGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTT 1020

Qy    1021 TCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTG 1080

Qy    1081 GCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCA 1140

Qy    1141 GCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 1200

Qy    1201 AAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 1260

Qy    1261 ACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAA 1320

Qy    1321 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTT 1380

Qy    1381 TCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAG 1440

Qy    1441 AAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 AAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAA 1500

Qy    1501 CGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1501 CGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTT 1560

Qy   1561 TAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCA 1620

Qy   1621 TTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGA 1680

Qy   1681 TTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTA 1740

Qy   1741 AAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAAC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAAC 1800

Qy   1801 AGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTG 1860

Qy   1861 TATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAA 1920

Qy   1921 TTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTA 1980

Qy   1981 GACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATT 2040

Qy   2041 AAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTTCAGCACCTGGGAAGTAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATTTTCAGCACCTGGGAAGTAA 2100

Qy   2101 TCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTTCTGCTACTGATGACACTCATTGT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTTCTGCTACTGATGACACTCATTGT 2160

Qy   2161 CATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAATGATACCTATAAATATATTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAATGATACCTATAAATATATTT 2220

Qy   2221 ATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATAAATGGTGTCATTTATGAACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGA 2280

Qy   2281 AGCCCTAATATTTAAAATGGCCTTATTTTACCATATGGATATAAGATTTGGCTCATAATG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AGCCCTAATATTTAAAATGGCCTTATTTTACCATATGGATATAAGATTTGGCTCATAATG 2340

Qy   2341 ATGAGCCCTATCATTTGATTTGAGTTCTATCATTTAAGAGAGCCTAAATAAAATTATCAT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 ATGAGCCCTATCATTTGATTTGAGTTCTATCATTTAAGAGAGCCTAAATAAAATTATCAT 2400

Qy   2401 CAAGGTATTAAATATAAGACGTTAAATATAATAAAGTGGGGATATATAGAAAACACACAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CAAGGTATTAAATATAAGACGTTAAATATAATAAAGTGGGGATATATAGAAAACACACAG 2460

Qy   2461 TGTTAGCACAGAGTAAGATCTCAATGCACATTTGTTGGATGAATAAATAAATGCAATTGA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TGTTAGCACAGAGTAAGATCTCAATGCACATTTGTTGGATGAATAAATAAATGCAATTGA 2520

Qy   2521 ATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAGTTCTACTTTAGAAGTACTAATTGG 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 ATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAGTTCTACTTTAGAAGTACTAATTGG 2580

Qy   2581 AGATGACTTTTATATCCCATTTTGGTAATTATTCATACATAGCACATATGACCATGATGT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGATGACTTTTATATCCCATTTTGGTAATTATTCATACATAGCACATATGACCATGATGT 2640
```

```
Qy   2641 TCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 2676
          |||||||||||||||||||||||||||||||||||
Db   2641 TCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 2676


RESULT 2
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag= a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 134-142; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
```

```
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

  Query Match           57.6%;  Score 1542;  DB 21;  Length 2822;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1542;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     287 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340

Qy     347 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400

Qy     407 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460

Qy     467 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520

Qy     527 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 586
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580

Qy     587 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640

Qy     647 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 706
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700

Qy     707 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760

Qy     767 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820

Qy     827 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880

Qy     887 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940

Qy     947 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000

Qy    1007 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060

Qy    1067 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120

Qy    1127 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180

Qy    1187 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240

Qy    1247 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300

Qy    1307 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360

Qy    1367 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 1426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2361 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 2420

Qy    1427 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 1486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2421 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 2480

Qy    1487 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 1546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2481 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 2540

Qy    1547 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 1606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2541 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 2600

Qy    1607 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 1666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2601 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 2660

Qy    1667 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 1726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2661 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 2720

Qy    1727 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 1786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2721 AGATGCACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACT 2780

Qy    1787 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 1828
           |||||||||||||||||||||||||||||||||||||||||||
Db    2781 TTGCATACTATAACAGAGGAAGAACATGTTGCGATTGAATTC 2822


RESULT 3
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag= a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
```

PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
DR P-PSDB; AAY80995.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Claim 4; Page 127-134; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC (AAY80995), and sequences AAZ92670-Z92673 are murine
CC alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC amplification of cDNA ends).
XX
SQ Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

Query Match 45.1%; Score 1206.8; DB 21; Length 2170;
Best Local Similarity 78.4%; Pred. No. 3.8e-260;
Matches 1674; Conservative 0; Mismatches 387; Indels 75; Gaps 16;

```
Qy   187 CCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATA 246
         | || ||| || |||||||||| |||||||| | || |||||  || ||| ||||||  |
Db     9 CGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGAGTAGTTCA 68

Qy   247 ACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACA 306
           ||    |||   ||||| ||||| |||||||||| | ||||||||||||||||||||
Db    69 GCACTCATCTC--TGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACA 126

Qy   307 TCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCA 366
         |||||||| |||||||||||||| |||| |||||||| |||||||||| |||||||||
Db   127 TCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCA 186

Qy   367 TGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCC 426
         ||||| |||||||| ||||| || |||||||||||| |||||||||||||| || ||
Db   187 TGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCA 246

Qy   427 AGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACT 486
         || |||||||||||||||||| ||||| ||||| |||||||||||||||| ||||||
Db   247 AGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACT 306

Qy   487 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 546
         || |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db   307 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 366

Qy   547 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
         |||||||| |||||||| |||||||||||||| ||||| ||| ||||||||||||||||
Db   367 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426

Qy   607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 666
         || || ||||| || ||  |||||||||| |||||||||| || ||||||||||| ||
Db   427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 486

Qy   667 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 726
         |||||||||| |||||||||||||||||||||||| || ||||| ||||||| ||||||
Db   487 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 546

Qy   727 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 786
         |||||||||||||| |||||||| |||||||||| ||||| ||||||||||||||||||
Db   547 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 606

Qy   787 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 846
         |||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||
Db   607 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 666

Qy   847 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 906
         || ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   667 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 726

Qy   907 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 966
         ||||||| ||||| ||||| || |||||||||| ||||| || |||||||| ||||||
Db   727 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 786

Qy   967 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 1026
         || || || |||||||||| |||||||| |||||||||| |||||||||||||||||| |||
Db   787 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 846

Qy  1027 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 1086
         ||||| ||||| ||||| |||||||||| ||||||| ||||||| ||| |||||||
Db   847 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 906

Qy  1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146
         || |||||||||| ||||| |||||||||||||||||||||||||||||||||||| |||
Db   907 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 966

Qy  1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 1206
         |||||||||||||||||||| |||||||||||||| ||||||| ||||||||||||
Db   967 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 1026

Qy  1207 GTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTA 1266
         || |||||||||||||||||| |||||||||||||||||||||||||| |||||||||||
Db  1027 GTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTA 1086

Qy  1267 AATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCAT 1326
         || || ||||| |||||||||||||||||||  |||| ||||| ||||| |||||||||
Db  1087 AACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCAT 1146

Qy  1327 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCA 1386
         |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| | | |||||
Db  1147 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT-AAAGTGTCCATCA 1205

Qy  1387 CTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGA 1446
           |  ||||   |  ||| |  || |||||| || || ||||| || ||||||||||
Db  1206 TTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAAGTTTTCAGGATAAGAAGA-- 1260

Qy  1447 TGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTA 1506
           ||||  || | |||| | |||  |||| ||||    ||||| || | | | | ||||
Db  1261 ---AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGCCCTCCCTTGAGGACCATGTA 1317

Qy  1507 TATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATT 1566
         ||||||||| || |||||||| || |||| || |||||||| || ||| ||||||||||
Db  1318 TATTTTGGT-GAAATTTTTAAGAGATCAGAATTAGCAATCACTCAATTTGGTTTTAAATT 1376

Qy  1567 ATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTA 1626
         |||||||||||  | || | | |||||||| | |||||||||  ||| | |||||| ||
Db  1377 ATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTTATTTGTCACTCTCATTTATA 1435

Qy  1627 AACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTTGTGGTTTGATTATTG 1686
         |||||| |||| ||   |||| |||||| |  ||| ||||  || || ||   |  |||
Db  1436 AACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTAAATTTATGATT---TACTTG 1489
```

```
Qy   1687 TTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATGCACATCTTAAAGTAT 1746
          |||| |||   ||||| | ||||||||||||||||| || || || || ||| |||| |||
Db   1490 TTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAAAAATACATATCCTAAAATAT 1549

Qy   1747 GA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCATACTATAACAGAGG 1805
          ||   | ||||  | |||||||| || || ||| | | |  ||| ||| |   |   | |
Db   1550 GAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAGCTTGTATAATGCTAACAAAG 1609

Qy   1806 AAGAA--CATGTTGCGATTGAATTCTAACCTCTTTGACTCCTAAGATGAATGAAGTGTAT 1863
          |||||    ||   |  || | |  ||| |||||||||| | ||| ||||       |
Db   1610 AAGAATTATTGAATCACTTAATTCTTAATCTCTTTGACTTCAAAGTTGAACACTACTGAC 1669

Qy   1864 AACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCACATTTGTGAAGGTGAAATTA 1923
          |||||             ||    |||     | || ||||||||||||| ||  | |
Db   1670 AACTG-------------ATAGAGTTTTACTTGTACCACATTTGTGAAGGGGAGCTAA 1715

Qy   1924 TTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTCTATTCAGGAAGCTATTAGAC 1983
          || ||| ||  |||||||||||||||||||||| |||              ||||||||
Db   1716 TTGATGTAGCAAATAAGAAAGATATGAAGCAGAAATGT-------------TATTAGAC 1761

Qy   1984 TTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTATTCTCATGGTCTTAAATTAAA 2043
          |||  |||| ||||| |||||| || || |||  |    |||| ||   | |||| |||||
Db   1762 TTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGATTCCCAATATTTTAAGTTAAA 1821

Qy   2044 TTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATT----------------TTCAGC 2087
            |  |||| ||||| | |||  |||||| |   | ||                ||
Db   1822 GAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTTAGGGAGATTTGATTGTTTGTT 1881

Qy   2088 ACCTGGGAAGTAATCCCAATAATACTTTAGAAAATC-TAAGACAGTTCTTTCTGCTACTG 2146
            ||||       || |||||||| | ||| |||| |  || |||||  | |   |
Db   1882 TTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTTTGATAGTTCCCTTTCAAATCA 1941

Qy   2147 ATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAAATAACAAAGAAAAA------ 2200
            | |       |  |   |||| |||||||||||||  |||||||  |||||
Db   1942 TGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCTTATAACAATAAAAAAATGATA 2001

Qy   2201 -TGATACCTATAAATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGTAT 2259
           |  || ||||| || |||||||  |  |||||||  |       | | | |   |
Db   2002 TCTTCAAATACAAATAAATGTATAAATTATACCATTTATAGAACGCTAGTTACT-TCATT 2060

Qy   2260 CAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAA 2295
          || || ||   |||||||  ||||    ||||||||
Db   2061 CATTTGAAATATTTTTTCCAAAGCATCCATATTTAA 2096
```

RESULT 4
AAZ92672
ID AAZ92672 standard; cDNA; 2036 BP.
XX
AC AAZ92672;
XX
DT 05-JUN-2000 (first entry)
XX
DE Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; murine; mouse; ss.
XX
OS Mus sp.
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Kudo T, Sasaki K;
XX
DR WPI; 2000-183120/16.
XX
PT Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT epitope and an antibody recognizing it useful for diagnosis of brain
PT and kidney disease and cancer. -
XX
PS Example 8; Page 162-164; 172pp; Japanese.
XX
CC The invention relates to a novel alpha-1,3-fucosyltransferase which
CC transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC (generating the Lewis x or y epitope). It does not transfer a fucose
CC moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC and therefore does not generate the sialyl-Lewis x epitope. The
CC invention also relates to DNA sequences encoding alpha-1,3-
CC fucosyltransferase and expression vectors and host cells comprising
CC these DNA sequences. The invention additionally encompasses the
CC preparation of alpha-1,3-fucosyltransferase via the culture of
CC transformed cells or by expression of the protein in a transgenic animal;
CC antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC screening potential inhibitors or potentiators of
CC alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC compounds having fucose-containing sugar chains by use of the protein;
CC and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC FUC-TIV and is expressed mainly in brain and kidney tissues.
CC Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC potentiators and inhibitors may be used for the treatment and diagnosis
CC of diseases of the brain and kidney, and of cancers. They may be used for
CC the identification of substances which affect the activity or expression
CC of alpha-1,3-fucosyltransferase; such substances may be used
CC therapeutically. The knockout animals can be used to study the mechanisms
CC of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC (AAY80995), and sequences AAZ92670-Z92673 are murine
CC alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC amplification of cDNA ends).
XX
SQ Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

Query Match 43.3%; Score 1159.2; DB 21; Length 2036;
Best Local Similarity 78.6%; Pred. No. 1.6e-249;
Matches 1600; Conservative 0; Mismatches 363; Indels 73; Gaps 15;

```
Qy    287 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 346
          |||||||||||||||||||||||||||||||| ||||||||||||||| |||| |||||||| |
Db      1 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 60

Qy    347 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 406
          |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db     61 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 120

Qy    407 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 466
          |||||||||||||  || || || |||||||||||||||||||| ||||| ||||| ||||
Db    121 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 180

Qy    467 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 526
          |||||||||||||| |||||||| |||||||| ||||| |||||||||||||||||||||||
Db    181 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 240

Qy    527 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 586
          |||||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |
Db    241 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 300

Qy    587 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 646
          || |||||||||||||||||| || ||||| || ||  |||||||||| ||||||||||
Db    301 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 360

Qy    647 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 706
          | || |||||||||| || |||||||||| |||||||||||||||||||||||| || |
Db    361 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 420

Qy    707 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 766
```

```
Db    421 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 480

Qy    767 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 826

Db    481 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 540

Qy    827 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 886

Db    541 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 600

Qy    887 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 946

Db    601 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 660

Qy    947 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 1006

Db    661 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 720

Qy   1007 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1066

Db    721 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 780

Qy   1067 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1126

Db    781 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 840

Qy   1127 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1186

Db    841 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 900

Qy   1187 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1246

Db    901 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 960

Qy   1247 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1306

Db    961 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1020

Qy   1307 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1366

Db   1021 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1080

Qy   1367 GGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAA 1426

Db   1081 GGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACATAATTCAA 1136

Qy   1427 GTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCAC 1486

Db   1137 GTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAATTTTCCTGC 1191

Qy   1487 CCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATC 1546

Db   1192 CCTCCCTTGAGGACCATGTATATTTTGGT-GAAATTTTTAAGAGATCAGAATTAGCAATC 1250

Qy   1547 ATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTT 1606

Db   1251 ACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA-AATAATTT 1309

Qy   1607 ATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTA 1666

Db   1310 ATTTGTCACTCTCATTTATAAACATTGTTTTTACA---TTTTATAGTTGACTGTAAAGTA 1366

Qy   1667 AGCTTGTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGA 1726

Db   1367 AATTTATGATT---TACTTGTTTCTACATCAATCAGATCTTTAATCTATTTGGGAGATAA 1423

Qy   1727 AGATGCACATCTTAAAGTATGA-AAAATTTTCACTAAGTATTACAATGTCTAGTTCCAAC 1785

Db   1424 AAATACATATCCTAAAATATGAGGGACTTTTTGCCAAGTATTATAAAGTATAGATTCTAG 1483

Qy   1786 TTTGCATACTATAACAGAGGAAGAACA--TGTTGCGATTGAATTCTAACCTCTTTGACTC 1843

Db   1484 CTTGTATAATGCTAACAAAGAAGAATTATTGAATCACTTAATTCTTAATCTCTTTGACTT 1543

Qy   1844 CTAAGATGAATGAAGTGTATAACTGTCTCTATTTGATCTATTTTTTTTACCTGTTTATCA 1903

Db   1544 CAAAGTTGAACACTACTGACAACTG-------------ATAGAGTTTTTACTTGTACCA 1589

Qy   1904 CATTTGTGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAACTGTTC 1963

Db   1590 CATTTGTGAAGGGGAGCTAATTGATGTAGCAAATAAGAAAGATATGAAGCAGAAATGT-- 1647

Qy   1964 TATTCAGGAAGCTATTAGACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCTACTTAT 2023

Db   1648 ------------TATTAGACTTCACATTCATTTTTATTAAGTTGTTTAGCAAATGTAGAT 1695

Qy   2024 TCTCATGGTCTTAAATTAAATTATTCAAGTATTTTTAAATATCCAATTTGTTGTGATT-- 2081

Db   1696 TCCCAATATTTTAAGTTAAAGAAAACAAGAATTTTCAGATAAACAATTTATATGGTTTTA 1755

Qy   2082 --------------TTCAGCACCTGGGAAGTAATCCCAATAATACTTTAGAAAATC-TAA 2126

Db   1756 GGGAGATTTGATTGTTTGTTTTGTGGGTTTAGTCCCTAATAATACCTAAGAGAATCTTTT 1815

Qy   2127 GACAGTTCTTTCTGCTACTGATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAA 2186

Db   1816 GATAGTTCCCTTTCAAATCATGAAGATAACCACTATTTTAAAAATAAATAATTTCCTCTT 1875

Qy   2187 ATAACAAAGAAAAA-------TGATACCTATAAATATATTTATAAATGGTGTCATTTATG 2239

Db   1876 ATAACAATAAAAAAATGATATCTTCAAATACAAATAAATGTATAAATTATACCATTTATA 1935

Qy   2240 AACAATGTTTAATTATGTATCAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAA 2295

Db   1936 GAACGCTAGTTACT-TCATTCATTTGAAATATTTTTTCCAAAGCATCCATATTTAA 1990
```

RESULT 5
AAT59505
ID AAT59505 standard; cDNA; 1814 BP.
XX
AC AAT59505;
XX
DT 06-MAY-1997 (first entry)
XX
DE Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW septic shock; septicaemia; therapy; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 325..1353
FT /*tag= a
XX
PN WO9640881-A1.
XX
PD 19-DEC-1996.
XX
PF 08-MAY-1996; 96WO-US06427.
XX
PR 07-JUN-1995; 95US-0483151.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Holgersson J, Seed B;
XX
DR WPI; 1997-108639/10.
DR P-PSDB; AAW11820.
XX
PT New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS Claim 6; Fig 3; 58pp; English.

```
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             4.8%;  Score 127.4;  DB 18;  Length 1814;
  Best Local Similarity   52.1%;  Pred. No. 1.2e-18;
  Matches  334;  Conservative    0;  Mismatches  301;  Indels    6;  Gaps    2;

Qy     563 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 622
           | |||| ||| |   |  ||||   |||     ||   |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     623 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 682
           ||||  ||  |        |   |   |||| | | |     |||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     683 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 742
             ||| | |||       |||||||  || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     743 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 802
           | | || |||  |   ||||  || || ||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     803 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 862
            ||||  |     | |    ||   |   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     863 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 922
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     923 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 982
           |    |  |  |  ||   | ||| |  ||   |||      |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     983 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1042
           | || | || ||  |  |    |     || || ||  ||||||| || |||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    1043 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1099
           ||| ||||||||| || ||| | |    ||||||  |   ||||||  |||||||  || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    1100 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1159
             ||||||||  || |||  | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    1160 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 1200
           |||| || |  | |||| || |||| || ||    |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag= a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag= b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag= c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag= d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag= e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag= f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag= g
FT                   /number= 3
FT   exon            2067..3552
FT                   /*tag= f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
```

```
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             4.7%;  Score 127;  DB 18;  Length 3594;
  Best Local Similarity   52.3%;  Pred. No. 1.9e-18;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     563 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 622
           | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336

Qy     623 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 682
           ||||  ||  |           |   |     |||| | | |    ||||||| |   |||
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     683 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 742
             ||| | |||       |||||||  || | | | || |          | || |   |
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     743 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 802
           | | || |||  |   ||||  || || ||||||||||||||||   || || || ||
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     803 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 862
           ||||  |     | |    ||   |   |    | | ||   ||||   |   |||
Db    2517 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     863 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 922
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     923 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 982
           |       |  |  |  ||   | ||| |  ||    |||        |    | |||
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     983 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1042
           | || | || ||  |  |     || || ||  |||||| || |||   ||
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy    1043 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1099
           ||| |||||||| || ||| | |    ||||||  |   |||||| |||||| || | |
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy    1100 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1159
             |||||||  || |||  | ||| |||  || | | ||| ||||| | ||  | || |
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy    1160 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
           |||| || |  | |||| || |||| || ||
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904


RESULT 7
AAQ77732
ID   AAQ77732 standard; cDNA to mRNA; 1701 BP.
XX
AC   AAQ77732;
XX
DT   22-JUN-1995  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; sialylated Lewis blood group antigen;
KW   Lewis x; Lewis a; glycosylation; glycoprotein; glycolipid; ds.
XX
OS   Homo sapiens (THP-1 cells).
XX
FH   Key             Location/Qualifiers
FT   CDS             123..1151
FT                   /*tag= a
FT                   /product= alpha-1,3-fucosyltransferase
XX
PN   WO9423021-A.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994;   94WO-JP00496.
XX
PR   29-MAR-1993;   93JP-0069016.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hanai N,  Kurata K,  Nishi T,  Sasaki K;
XX
DR   WPI; 1994-333183/41.
DR   P-PSDB; AAR63215.
XX
PT   Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
PT   of active sugar chains, e.g. sialylated Lewis X
XX
PS   Claim 3; Page 99-102; 155pp; Japanese.
XX
CC   The sequence coding for alpha-1,3-fucosyltransferase was isolated
CC   from a human THP-1 cell cDNA library. The enzyme catalyses the
CC   addition of fucose via a 1,3-linkage to N-acetylglucosamine
CC   contained in glycoproteins and glycolipids, e.g. for sialylation of
CC   Lewis blood group antigens.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             4.3%;  Score 114.4;  DB 15;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 9.4e-16;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     550 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 609
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     610 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 669
           ||  |  |||| ||||| ||  |           | | |  |  | || | |      ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     670 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 729
           || |     |||   ||| | |||        |||| || || |  |||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     730 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 789
           ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     790 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 849
           |||||  ||||||  | ||  |     |      ||||    |      |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     850 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 909
             |   |||    |||||| || || ||||   |      |||     | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     910 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 969
           |||    ||||            |     |  ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     970 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1029
           |    |  |     || | || ||| |  |     |    || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    1030 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1086
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869
```

```
Qy   1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146
          |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy   1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
          | ||| | |||||||| || | |  |||  || | ||||| ||
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 8
AAF21136
ID AAF21136 standard; DNA; 1701 BP.
XX
AC AAF21136;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2703.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 1701 BP; 290 A; 527 C; 589 G; 295 T; 0 other;

Query Match 4.3%; Score 114.4; DB 21; Length 1701;
Best Local Similarity 50.8%; Pred. No. 9.4e-16;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy    550 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 609
          | |  ||||   | ||||| || |   |  ||||   |||     ||   || | ||
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    610 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 669
          ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    670 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 729
          || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    730 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 789
          ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    790 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 849
          |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    850 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 909
            |   |||     |||||| || || ||||   |      |||     | | || |
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    910 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 969
          |||    ||||           |      |  ||  ||  || |  ||   ||||
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    970 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1029
          |   |  |      || | || ||| |  | - |      || || || |||||||||
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy   1030 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1086
          || ||    |||   || ||||| ||||| ||  | |     ||| ||  |  ||||||||
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy   1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146
          |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy   1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
          | ||| | |||||||| || | |  |||  || | ||||| ||
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 9
AAA35014
ID AAA35014 standard; DNA; 1701 BP.
XX
AC AAA35014;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2703.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;

KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 957-958; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

Query Match 4.3%; Score 114.4; DB 21; Length 1701;
Best Local Similarity 50.8%; Pred. No. 9.4e-16;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy    550 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 609
          | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    610 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 669
          ||  |  |||| ||||| ||  |         | | |  |  | || | |       ||
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    670 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 729
          || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    730 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 789
          ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    790 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 849
          |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    850 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 909
            |   |||    |||||| || || ||||   |      |||     | | || |
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    910 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 969
          |||    ||||           |     |  ||  ||  || |  ||   ||||
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    970 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1029
          |   |  |      || | || ||| |  |     |     || || || ||||||||
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    1030 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1086
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146
           |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 10
AAF21137
ID AAF21137 standard; DNA; 6944 BP.
XX
AC AAF21137;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2704.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not

PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034-1036; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

Query Match 4.3%; Score 114.4; DB 21; Length 6944;
Best Local Similarity 50.8%; Pred. No. 1.6e-15;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy    550 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 609
          | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db   5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy    610 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 669
          ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db   5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy    670 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 729
          || |   |||   ||| | |||       |||| || || |  ||||| | |
Db   5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy    730 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 789
          ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db   5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy    790 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 849
          ||||| ||||||  | ||  |     |       ||||    |       |||||  |||
Db   5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy    850 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 909
            |  |||    |||||| || || ||||   |      |||     | | || |
Db   5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy    910 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 969
          |||    ||||           |      |  ||  ||  || |  ||   ||||
Db   5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy    970 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1029
          |   |  |      || | || ||| |  |     |      || || || ||||||||
Db   5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy   1030 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1086
          || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db   6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy   1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146
          |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db   6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy   1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
          | ||| | |||||||| || | |  |||  || | ||||| ||
Db   6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

RESULT 11
AAA35015
ID AAA35015 standard; DNA; 6944 BP.
XX
AC AAA35015;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2704.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 958-960; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present

```
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

  Query Match             4.3%;  Score 114.4;  DB 21;  Length 6944;
  Best Local Similarity   50.8%;  Pred. No. 1.6e-15;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     550 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 609
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     610 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 669
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     670 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 729
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     730 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 789
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     790 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 849
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     850 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 909
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     910 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 969
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     970 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1029
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy    1030 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1086
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy    1087 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1146
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy    1147 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216


RESULT 12
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag= a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             4.2%;  Score 113.4;  DB 15;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.6e-15;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
```

```
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
          |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
            ||| ||| ||||  | ||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1252 AAG 1254
            ||
Db   1095 GAG 1097


RESULT 13
AAT61680
ID   AAT61680 standard; DNA; 1654 BP.
XX
AC   AAT61680;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..1208
FT                   /*tag=  a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13643.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 6; Page 291-292; 329pp; English.
XX
CC   A DNA clone (AAT61680) codes for an enzyme (AAW13643) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC   precursors but not type I glycoconjugates.  It was identified in a
CC   human lambda phage genomic DNA library using a probe obtd. from
CC   Lewis enzyme cDNA (see also AAT61675).  The Fuc-TVI DNA can be used
CC   to construct animal cell lines with specific capabilities with
CC   respect to post-translational modification of the oligosaccharides
CC   of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC   use in oligosaccharide prodn., to genotype individuals at this
CC   fucosyltransferase locus, and in the prodn. of chimeric
CC   fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match              4.2%;  Score 113.4;  DB 18;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.6e-15;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
          || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
          |       ||      |  ||| |  ||||  | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
          || ||  |  | || |||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
          | ||||  |      |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
          ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
          |||      ||| || ||||           | | | |       |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
          |  |  |   |||      |  |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
          ||||| ||    || ||| |   ||           |  |  |  |   |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
            |     |    |     |   ||   ||||    ||  | ||   |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
          |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
            ||| ||| ||||  | ||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1252 AAG 1254
            ||
Db   1095 GAG 1097


RESULT 14
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
```

DE Lewis blood group fucosyltransferase DNA.
XX
KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW surface; oligosaccharide; ss.
XX
PN WO9402616-A.
XX
PD 03-FEB-1994.
XX
PF 20-JUL-1993; 93WO-US06703.
XX
PR 20-JUL-1992; 92US-0914281.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Lowe JB;
XX
DR WPI; 1994-048874/06.
XX
PT DNA fragment encoding a glycosyltransferase - can be used for in
PT vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT blood gp. determinants, to protect against transplant rejection
XX
PS Disclosure; Fig 6; 249pp; English.
XX
CC The sequence is that encoding a Lewis blood group fucosyl
CC transferase (Fuc-TIII). The enzyme produced by the DNA may be
CC non glycosylated. This prevents premature loss of enzyme activity.
CC It can also be used in in vitro reactions to modify cell surface
CC oligosaccharide mols. e.g. blood group determinants.
CC See also AAQ56905-11.
XX
SQ Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

Query Match 4.2%; Score 112.2; DB 15; Length 1086;
Best Local Similarity 51.6%; Pred. No. 2.4e-15;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy    565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
          | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
              |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
          |   ||  |   | |   |   |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
          |   ||           |  |  |  |    |||||  |   |               |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807
```

```
Qy   1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
          || || || || ||||| ||    ||    |||||  ||||  |   | |||  | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972
```

RESULT 15
AAT80111
ID AAT80111 standard; cDNA; 2042 BP.
XX
AC AAT80111;
XX
DT 01-DEC-1997 (first entry)
XX
DE Fucosyltransferase III coding sequence.
XX
KW P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW adult respiratory distress syndrome; tumour metastasis;
KW fucosyltransferase III; acetylglucosaminyltransferase;
KW autoimmune disorder; inflammatory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO9706176-A2.
XX
PD 20-FEB-1997.
XX
PF 02-AUG-1996; 96WO-US12820.
XX
PR 17-MAY-1996; 96US-0649802.
PR 03-AUG-1995; 95US-0510920.
PR 15-MAY-1996; 96US-0017794.
XX
PA (OKLA) UNIV OKLAHOMA STATE.
XX
PI Cummings RD, Mcever RP, Moore KL;
XX
DR WPI; 1997-154206/14.
XX
PT New O-glycan(s) derived from P-selectin glyco:protein ligand
PT (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT for treating inflammatory or autoimmune disorders or tumours
XX
PS Disclosure; Page 80-81; 96pp; English.
XX
CC This sequence represents the coding sequence for the human
CC fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC protein expressed on the surfaces of activated platelets and endothelium
CC in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC high affinity P-selectin ligand produced by leukocytes. Binding of
CC P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC highly extended extracellular domain, which is a feature of mucin-like
CC proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC O-glycans, and O-glycanated products of the invention. The products are
CC used for inhibiting the binding of P-selectin and other selectins to
CC PSGL-1. They can be used for the treatment of inflammation, injury

CC resulting from ischaemia and reperfusion, bacterial sepsis and
CC disseminated intravascular coagulation, adult respiratory distress
CC syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC other autoimmune or inflammatory disorders. The products can also be used
CC in the production of antibodies and in diagnostic applications.
XX
SQ Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

Query Match 4.2%; Score 112.2; DB 18; Length 2042;
Best Local Similarity 51.6%; Pred. No. 3.1e-15;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy     565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
           ||||| |||| | |||||     |||||     |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
           | || |||| |   |   |  | ||    | || |       | ||||| |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
           ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
           |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
           |   ||  |    |  |    |     |          | |  |  |  |   |||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
           |    ||          |  |  |  |    ||||| |  |     |              |
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
           || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
           ||  | |||   || |||||||| || |||||  |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
           || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
           ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
           |||||  || |       ||||| || ||||||  |||| || ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 5, 2002, 23:15:38
Job time : 369.303 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 2240.54 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-5
Perfect score: 2676
Sequence: 1 gcgcgcgcggcgcagcagct.....ccaaataaacctaccattac 2676

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		603.2	22.5	716	9	AV725071	AV725071 AV725071
2		575.6	21.5	697	9	BB397507	BB397507 BB397507
3	c	573.2	21.4	579	10	BM021081	BM021081 ie75f03.x
4		532.2	19.9	568	10	BM021317	BM021317 ie75f03.y
5		492.8	18.4	607	10	BI102760	BI102760 602888338
6		480.4	18.0	490	12	AQ489567	AQ489567 RPCI-11-2
7	c	448	16.7	501	12	B41073	B41073 HS-1052-B2-
8		430	16.1	649	9	AV338075	AV338075 AV338075
9	c	411.6	15.4	548	12	AQ772584	AQ772584 HS_5568_A
10		390.6	14.6	648	9	BB595164	BB595164 BB595164
11		378.4	14.1	625	9	BB649967	BB649967 BB649967
12		366.6	13.7	767	9	BB640911	BB640911 BB640911
13	c	357	13.3	646	12	AQ418233	AQ418233 RPCI-11-1
14		325.6	12.2	628	9	BB651409	BB651409 BB651409
15	c	296.8	11.1	1101	12	CNS059ST	AL327638 Tetraodon
16		242.6	9.1	670	9	BB626126	BB626126 BB626126
17		240.2	9.0	533	10	BI604066	BI604066 603243886
18	c	197.4	7.4	335	12	AQ054539	AQ054539 CIT-HSP-2
19		197	7.4	617	10	BE266935	BE266935 601189956
20	c	193.2	7.2	487	12	AQ246164	AQ246164 HS_2060_A
21	c	184.2	6.9	570	10	BM082484	BM082484 fu25e08.x
22	c	177	6.6	1100	12	CNS0579Z	AL324368 Tetraodon
23	c	170.2	6.4	1016	12	CNS0429P	AL271222 Tetraodon
24		165	6.2	844	12	CNS01V8V	AL168808 Tetraodon
25		163	6.1	650	9	AV375511	AV375511 AV375511
26	c	157.4	5.9	887	12	CNS02Q0S	AL208693 Tetraodon
27		155.4	5.8	651	10	BJ036855	BJ036855 BJ036855
28		153	5.7	367	9	BB843869	BB843869 BB843869
29	c	151.6	5.7	666	10	BJ050300	BJ050300 BJ050300
30	c	151.2	5.7	711	12	CNS034XF	AL228012 Tetraodon
31		146.6	5.5	640	10	BJ073227	BJ073227 BJ073227
32		138	5.2	199	9	BB398103	BB398103 BB398103
33	c	125.2	4.7	674	10	BE605503	BE605503 fl10h04.x
34		120.6	4.5	971	12	CNS047IV	AL278032 Tetraodon
35	c	120.4	4.5	615	10	BI326811	BI326811 STR00078
36		119.8	4.5	284	9	BB270301	BB270301 BB270301
37	c	117.8	4.4	1023	12	CNS02ANG	AL188773 Tetraodon
38		115.6	4.3	317	9	BB272545	BB272545 BB272545
39	c	115.4	4.3	437	12	AQ819773	AQ819773 HS_5513_A
40		114.8	4.3	661	9	BB651647	BB651647 BB651647
41		106.2	4.0	347	10	M78549	M78549 EST00697 Fe
42	c	103	3.8	702	10	BJ043578	BJ043578 BJ043578
43		102.8	3.8	487	10	BE863396	BE863396 UI-M-BH0-
44		93.8	3.5	863	12	CNS02ROF	AL210840 Tetraodon
45		90.4	3.4	720	10	BF180380	BF180380 601804564

ALIGNMENTS

RESULT 1
AV725071
LOCUS AV725071 716 bp mRNA linear EST 16-OCT-2000
DEFINITION AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION AV725071
VERSION AV725071.1 GI:10830091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 716)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA HTC clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1..716
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBLB02"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 219 a 113 c 133 g 249 t 2 others
ORIGIN

```
  Query Match             22.5%;  Score 603.2;  DB 9;  Length 716;
  Best Local Similarity   99.2%;  Pred. No. 7.1e-103;
  Matches  605;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     953 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 1012
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy    1013 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 1072
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy    1073 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 1132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy    1133 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 1192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy    1193 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 1252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy    1253 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 1312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy    1313 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy    1373 AAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTG 1432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATTG 480

Qy    1433 AGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTC 1492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCTC 540

Qy    1493 TAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCA 1552
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     541 TAGGGTAACGTGTATATCTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCCA 600

Qy    1553 TTCGGTTTTA 1562
           ||    ||||
Db     601 TTNNCGTTTA 610
```

```
RESULT 2
BB397507
LOCUS       BB397507                 697 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA
            clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for
            alphal,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   BB397507
VERSION     BB397507.2  GI:16411808
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 15, 2000 this sequence version replaced gi:9216903.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .697
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="C330004C19"
                      /clone_lib="RIKEN full-length enriched, ES cells"
                      /cell_type="ES cells"
                      /lab_host="SOLR"
                      /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                      cDNA was prepared with the primer adapter of sequence [5'
                      GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN
```

```
  Query Match             21.5%;  Score 575.6;  DB 9;  Length 697;
  Best Local Similarity   91.2%;  Pred. No. 1e-97;
  Matches  622;  Conservative    0;  Mismatches   59;  Indels    1;  Gaps    1;

Qy     698 ATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGA 757
           | || || ||||| |||||||| ||| |||||||||||| ||| |||||||| |||||||
Db       4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy     758 CTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCA 817
```

```
Db     63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy    818 CAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTA 877
          ||||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||
Db    123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy    878 ACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAA 937
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||
Db    183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy    938 TCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCA 997
          |||| ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||
Db    243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy    998 TATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGG 1057
          ||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||| |
Db    303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy   1058 AAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGG 1117
          ||||||| |||||||| ||| ||||||| || |||||||||||| ||||| ||||||||||
Db    363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy   1118 AAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTC 1177
          |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||
Db    423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy   1178 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTT 1237
          |||||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db    483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy   1238 ACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCAT 1297
          |||||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||
Db    543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy   1298 GTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 1357
          |  |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db    603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy   1358 AATGGTTTTGGAATTAAAATTT 1379
          |||||||||||||||||| | |
Db    663 AATGGTTTTGGAATTAAAGTGT 684
```

```
RESULT 3
BM021081/c
LOCUS       BM021081                579 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021081
VERSION     BM021081.1  GI:16535437
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
BASE COUNT      203 a    103 c     93 g    180 t
ORIGIN
```

```
  Query Match             21.4%;  Score 573.2;  DB 10;  Length 579;
  Best Local Similarity   99.5%;  Pred. No. 2.9e-97;
  Matches  575;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   1066 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 1125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy   1126 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 1185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy   1186 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 1245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy   1246 TGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCT 1305
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy   1306 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy   1366 TGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 TGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCA 220

Qy   1426 AGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCA 1485
          ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 AGTATTGAAAATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCA 160

Qy   1486 CCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAAT 1545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 CCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAAT 100

Qy   1546 CATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATT 1605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 CATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATT 40

Qy   1606 TATTCTTCATTATCATTTGTAAACATTGCTTTTTCACA 1643
```

```
          ||||||||||||||||||||||||||||||||||||
Db      39 TATTCTTCATTATCATTTGTAAACATTGCTTTTTCACA 2


RESULT 4
BM021317
LOCUS       BM021317                 568 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021317
VERSION     BM021317.1  GI:16535673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ie75f03.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT      127 a    170 c    123 g    148 t
ORIGIN

  Query Match             19.9%;  Score 532.2;  DB 10;  Length 568;
  Best Local Similarity   99.4%;  Pred. No. 1.4e-89;
  Matches  534;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     32 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCG 91

Qy     61 CCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGCCGCCAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 CCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGCCGCCAC 151

Qy    121 CGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGGATGGCG 180
          |||||||||||||||||||||||||||||||||||||||||||||||| |  ||||||||
Db    152 CGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGTCTTGGATGGCG 211

Qy    181 CTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 CTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGT 271

Qy    241 AGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 AGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACA 331

Qy    301 TCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 TCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTC 391

Qy    361 ATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    392 ATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAA 451

Qy    421 TCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    452 TCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAAT 511

Qy    481 GAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    512 GAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 568


RESULT 5
BI102760
LOCUS       BI102760                 607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: 1  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
```

```
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN

  Query Match             18.4%;  Score 492.8;  DB 10;  Length 607;
  Best Local Similarity   90.1%;  Pred. No. 3.2e-82;
  Matches  549;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy     517 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACA 576
           |||||||||||| ||||||||||||||||||||||||||| |||||||| ||||||||||||
Db       1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy     577 ACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGT 636
           || ||||| ||| |||||||||||||||||| || ||||| || ||  ||||||||||
Db      61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy     637 TGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATG 696
           |||||||||| || |||||||||| || |||||||||| |||||||||| |||||||
Db     120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy     697 AATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTG 756
           ||||| || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||
Db     179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy     757 ACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGC 816
           ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||| |||
Db     239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy     817 ACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGT 876
           |||||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||
Db     299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy     877 AACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAA 936
           |||||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||
Db     359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy     937 ATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACC 996
           ||||| ||||| || |||||||| |||||||| || || ||||||||| ||||||| |||
Db     419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy     997 ATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACG 1056
           |||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||
Db     479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy    1057 GAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGG 1116
           |||||||| |||||||| ||| ||||||| || ||||||  ||| ||||| |||||||||
Db     539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy    1117 GAAAACTAT 1125
           |||||||||
Db     599 GAAAACTAT 607


RESULT 6
AQ489567
LOCUS       AQ489567                490 bp    DNA     linear   GSS 24-APR-1999
DEFINITION  RPCI-11-230K15.TV RPCI-11 Homo sapiens genomic clone RPCI-11-230K15
            , DNA sequence.
ACCESSION   AQ489567
VERSION     AQ489567.1  GI:4675441
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /db_xref="GDB:7588190"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-230K15"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      169 a     66 c     68 g    187 t
ORIGIN

  Query Match             18.0%;  Score 480.4;  DB 12;  Length 490;
  Best Local Similarity   98.8%;  Pred. No. 6.7e-80;
  Matches  484;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    1838 TGACTCCTAAGATGAATGAAGTGTATAACTGTCTCTATTTGATCTATTTTTTTTACCTGT 1897
           ||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1 TGACTCCGAAGATGAATGAAGTGTATAACTGTCTCTATTTGATTTATTTTTTTTACCTGT 60

Qy    1898 TTATCACATTTGTGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAA 1957
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTATCACATTTGTGAAGGTGAAATTATTCATGGAGTGAATAAGAAAGATATGAAGCAGAA 120

Qy    1958 CTGTTCTATTCAGGAAGCTATTAGACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCT 2017
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTGTTCTATTCGGGAAGCTATTAGACTTCTCATTTATTTTCATTAAGCTGATTTGCAGCT 180

Qy    2018 ACTTATTCTCATGGTCTTAAATTAAATTATTCAAGTATTTTTAAATATCCAATTTGTTGT 2077
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACTTATTCTCATGGTCTTAAATTAAATTATTCAAGTATTTTTAAATATCCAATTTGTTGT 240

Qy    2078 GATTTTCAGCACCTGGGAAGTAATCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTT 2137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GATTTTCAGCACCTGGGAAGTAATCCCAATAATACTTTAGAAAATCTAAGACAGTTCTTT 300

Qy    2138 CTGCTACTGATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAAATAACAAAGAA 2197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTGCTACTGATGACACTCATTGTCATAATAAAACAAATAATTTCCTCAAATAACAAAGAA 360

Qy    2198 AAATGATACCTATAAATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGT 2257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAATGATACCTATAAATATATTTATAAATGGTGTCATTTATGAACAATGTTTAATTATGT 420

Qy    2258 ATCAATTTAAGATTTTTTTCTGAAGCCCTAATATTTAAAATGGCCTTATTTTACCATATG 2317
           |||||||||||| |||||||||||||||||||||||||||| |||| ||||||||||
Db     421 ATCAATTTAAGAATTTTTTCTGAAGCCCTAATATTTAAAATGGTCTTATCTTACCATATG 480

Qy    2318 GATATAAGAT 2327
           ||||||||||
Db     481 GATATAAGAT 490


RESULT 7
B41073/c
LOCUS       B41073                  501 bp    DNA     linear   GSS 18-OCT-1997
DEFINITION  HS-1052-B2-G01-MF.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 774 Col=2 Row=N, DNA sequence.
ACCESSION   B41073
```

```
VERSION     B41073.1  GI:2545325
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 774  row: N  column: 2
            Class: BAC ends
            High quality sequence stop: 501.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=CT 774 Col=2 Row=N"
                     /clone_lib="CIT Human Genomic Sperm Library C"
                     /sex="M"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      179 a     95 c     83 g    143 t      1 others
ORIGIN

  Query Match             16.7%;  Score 448;  DB 12;  Length 501;
  Best Local Similarity   99.8%;  Pred. No. 7.7e-74;
  Matches  448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1224 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 1283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 442

Qy   1284 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 1343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 382

Qy   1344 TGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAA 1403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 TGGTAATTTAGAGAAATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAA 322

Qy   1404 TATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTT 1463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 TATTTTGATGAGATATCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTT 262

Qy   1464 GTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTT 1523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 GTGTCACAATTTATTTTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTT 202

Qy   1524 TTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAA 1583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TTAAAAGCTCAGCATGAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAA 142

Qy   1584 TTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACA 1643
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    141 TTATGTGCACTGGAGAGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTNCACA 82

Qy   1644 TTTTTGTAGTTGTCCATAATGTAAGCTTG 1672
          |||||||||||||||||||||||||||||
Db     81 TTTTTGTAGTTGTCCATAATGTAAGCTTG 53


RESULT 8
AV338075
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430407H24"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory bulb"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
```

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 209 a 105 c 117 g 218 t
ORIGIN

Query Match 16.1%; Score 430; DB 9; Length 649;
Best Local Similarity 83.8%; Pred. No. 1.8e-70;
Matches 548; Conservative 0; Mismatches 95; Indels 11; Gaps 5;

```
Qy    999 ATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGA 1058
          |||  ||||||| |||  |||||| |||||||| ||||| ||||||||||||||||| ||
Db      2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61

Qy   1059 AAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGA 1118
          ||||||  |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy   1119 AAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCC 1178
          |||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy   1179 CAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTA 1238
          |||||||  |||||||  |||||||||||||| |||||||||||||||| ||||||||||
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy   1239 CTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATG 1298
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy   1299 TTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 1358
            |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy   1359 ATGGTTTTGGAATTAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATA 1418
          |||||||||||||| ||| | | ||||| |  ||||   |  ||| |  || ||||||  |
Db    362 ATGGTTTTGGAATT-AAAGTGTCCATCATTGTCACA---GTGAAAAAGATTAATGAGACA 417

Qy   1419 TCATCCAAGTATTGAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATT 1478
          | || ||||| || ||||||||||||     ||||  || | |||| | |||  |||| |
Db    418 TAATTCAAGTTTTCAGGATAAGAAGA-----AACACGCTGCATTTGGGGCACCGTTTAAT 472

Qy   1479 TTTATCACCCTCTCTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCAT 1538
          |||    ||||| || | | | | |||||||||||||| || |||||||| || |||| ||
Db    473 TTTCCTGCCCTCCCTTGAGGACCATGTATATTTTGGT-GAAATTTTTAAGAGATCAGAAT 531

Qy   1539 GAGCAATCATTCCATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAG 1598
          ||||||||  || ||| |||||||||||||||||||||  | || | | ||||||||
Db    532 TAGCAATCACTCAATTTGGTTTTAAATTATCCTGTATATATGTGATAACGAGCACTGGA- 590

Qy   1599 AGTAATTTATTCTTCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAG 1652
          | |||||||||  ||| | |||||| |||||||| |||| ||  || | || |
Db    591 AATAATTTATTTGTCACTCTCATTTATAAACATTGTTTTTACATTTTATAGTTG 644
```

RESULT 9
AQ772584/c
LOCUS AQ772584 548 bp DNA linear GSS 29-JUL-1999

DEFINITION HS_5568_A2_D09_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1144 Col=18 Row=G, DNA sequence.
ACCESSION AQ772584
VERSION AQ772584.1 GI:5652312
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 548)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 1144 row: G column: 18
Seq primer: T7
Class: BAC ends
High quality sequence stop: 548.
FEATURES Location/Qualifiers
source 1. .548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1144 Col=18 Row=G"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 190 a 94 c 89 g 168 t 7 others
ORIGIN

Query Match 15.4%; Score 411.6; DB 12; Length 548;
Best Local Similarity 92.1%; Pred. No. 5.1e-67;
Matches 432; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy   1372 TAAAATTTTTCATCACTTGCACACTTGATAAATATTTTGATGAGATATCATCCAAGTATT 1431
          |||| |||||||||||| ||||||||||||||||||||| ||||||||||||||||||
Db    547 TAAATTTTTTCATCACTGGCACACTTGATAAATATTTTGATCAGATATCATCCAAGTATT 488

Qy   1432 GAGGATAAGAAGAGATGCAACATACTACTTTTGTGTCACAATTTATTTTTATCACCCTCT 1491
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    487 GAGGATAAGAAGAGATGCAACATCCTACTTTTGTGTCACAATTTATTTTTATCACCCTCT 428

Qy   1492 CTAGGGTAACGTGTATATTTTGGTGGAGATTTTTAAAAGCTCAGCATGAGCAATCATTCC 1551
          ||||||||||||||||||||||||||| |||||| |||||||||| | ||||||||||||
Db    427 CTAGGGTAACGTGTATATTTTGGTGGAAATTTTTTAAAGCTCAGCTTAAGCAATCATTCC 368

Qy   1552 ATTCGGTTTTAAATTATCCTGTATATACCTAATTATGTGCACTGGAGAGTAATTTATTCT 1611
          ||||||||||||| |||| ||||||||  |||||||||||||||||||||||||||||| |
Db    367 ATTCGGTTTTAAATGATCCGGTATATACGCAATTATGTGCACTGGAGAGTAATTTATTTT 308

Qy   1612 TCATTATCATTTGTAAACATTGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTT 1671
          |||||||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||
Db    307 TCATTATCATATGTAAGCATAGCTTTTTCACATTTTTGTAGTTGTCCATAATGTAAGCTT 248
```

```
Qy   1672 GTGGTTTGATTATTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAGATG 1731
          ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||| |||
Db    247 GTGGTTTGATACTTGTTTCCACACTGATCAGCTGTTTAATCTATTTGGGAAATGAAAATG 188

Qy   1732 CACATCTTAAAGTATGAAAAATTTTCACTAAGTATTACAATGTCTAGTTCCAACTTTGCA 1791
          ||||||||||||||| | |||||  | ||||||||||||||| ||||||||||||||||
Db    187 CACATCTTAAAGTATGTAGAATTTAAAATAAGTATTACAATGTGTAGTTCCAACTTTGCA 128

Qy   1792 TACTATAACAGAGGAAGAACATGTTGCGATTGAATTCTAACCTCTTTGA 1840
          |||||| ||| || |||||||||||| ||| |||| |  | | | || |
Db    127 TACTATGACANAGNAAGAACATGTTGNGATGGAATGCCCAGCGCCTTCA 79
```

RESULT 10
BB595164
LOCUS BB595164 648 bp mRNA linear EST 26-OCT-2001
DEFINITION BB595164 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION BB595164
VERSION BB595164.2 GI:16450167
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 648)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Nov 30, 2000 this sequence version replaced gi:11491766.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES Location/Qualifiers
source 1. .648
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030004N01"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 141 a 206 c 138 g 163 t
ORIGIN

Query Match 14.6%; Score 390.6; DB 9; Length 648;
Best Local Similarity 79.0%; Pred. No. 4.3e-63;
Matches 505; Conservative 0; Mismatches 124; Indels 10; Gaps 3;

```
Qy      1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCT------CCCCTGCAGCTCCCCGCG 54
          ||||||| ||||| |||||||  ||  |   | |||        ||| |   ||     |
Db     14 GCGCGCGGGGCGCTGCAGCTCTGGAGCCCTGGATCTTTCCTCACCCGGTGCCTTGAGACC 73

Qy     55 CCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGC 114
          || |||||||| ||||  | | | |  ||||   | || |   ||||  ||||||||||
Db     74 CCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTACGCTACTAGGACAGCGC 133

Qy    115 CGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCCGCAGCCCTCGCGAGCGCCCCGG 174
          ||||||| ||||| |  ||  | | || || |||||||| |||||| || || |
Db    134 CGCCACCACCGCCAG--CCCTCGTCCCGCCCGCGCCGCGCTGCCCTCCCGCGCCCTGGAT 191

Qy    175 ATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAG 234
             ||  |    | || ||| || |||||||||| |||||||| | || |||||  || |
Db    192 GGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCG 251

Qy    235 TGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATT 294
          || ||||||  | ||   |||   ||||| ||||| |||||||||| | |||||||||
Db    252 TGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTTCTCTACACTGAAAAATT 309

Qy    295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354
          |||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db    310 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 369

Qy    355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414
          || |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db    370 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 429

Qy    415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474
          ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db    430 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 489

Qy    475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534
          ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db    490 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 549
```

```
Qy      535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594
            ||||||||||||||||||| |||| ||| |||||||||||||| ||||| ||| |||||
Db      550 TCCTGCCAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 609

Qy      595 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATC 633
            ||||||||||||| || ||||| || ||   |||||||
Db      610 AACAAATCCCATGCGGTCCTGATTCACCATATAGACATC 648


RESULT 11
BB649967
LOCUS       BB649967                  625 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB649967 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION   BB649967
VERSION     BB649967.1  GI:16484222
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES
     source         Location/Qualifiers
                    1. .625
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="C130099N17"
                    /clone_lib="RIKEN full-length enriched, 16 days embryo
                    head"
                    /sex="mixed"
                    /tissue_type="head"
                    /dev_stage="16 days embryo"
                    /lab_host="DH10B"
                    /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN. Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer [5'
                    GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. Second strand cDNA was prepared with the
                    primer adapter of sequence [5'
                    GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                    was cloned into the XhoI and BamHI sites. Vector: a
                    modified pBluescript KS(+) after bulk excision from Lambda
                    FLC I"
BASE COUNT      136 a    198 c    135 g    156 t
ORIGIN


  Query Match             14.1%;  Score 378.4;  DB 9;  Length 625;
  Best Local Similarity   79.5%;  Pred. No. 8.3e-61;
  Matches  488;  Conservative    0;  Mismatches  116;  Indels   10;  Gaps    3;

Qy        1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCT------CCCCTGCAGCTCCCCGCG 54
            ||||||| ||||| |||||||  ||  |   | |||       ||| |  ||    |
Db       15 GCGCGCGGGGCGCTGCAGCTCTGGAGCCCTGGATCTTTCCTCACCCGGTGCCTTGAGACC 74

Qy       55 CCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGC 114
            || ||||||||  ||||  | | | |  ||||   | || |   ||||  ||||||||||
Db       75 CCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTACGCTACTAGGACAGCGC 134

Qy      115 CGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGG 174
            |||||||  |||||  |  ||  | | ||  || |||||||||  ||||||  || || |
Db      135 CGCCACCACCGCCAG--CCCTCGTCCCGCCCGCGCCGCGCTGCCCTCCCGCGCCCTGGAT 192

Qy      175 ATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAG 234
              ||  |    | || ||| || |||||||||  |||||||| | || |||||  || |
Db      193 GGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCG 252

Qy      235 TGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATT 294
            || ||||||  | ||    |||  ||||| ||||| ||||||||||  | |||||||||
Db      253 TGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTTCTCTACACTGAAAAATT 310

Qy      295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354
            |||||||||||||||||||| ||||||||||||||| |||| |||||||| |||||||||
Db      311 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 370

Qy      355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414
            || |||||||||||||| |||||||| ||||| || ||||||||||| ||||||||||
Db      371 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 430

Qy      415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474
            ||||| || || || |||||||||||||||||||| ||||| ||||| ||||||||||||
Db      431 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 490

Qy      475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534
            ||||| |||||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db      491 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 550

Qy      535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594
            |||||||||||||||||||| |||||||| |||||||||||||| ||||| ||| |||||
Db      551 TCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTAC 610

Qy      595 AACAAATCCCATGC 608
            ||||||||||||||
Db      611 AACAAATCCCATGC 624
```

RESULT 12
BB640911
LOCUS BB640911 767 bp mRNA linear EST 26-OCT-2001
DEFINITION BB640911 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730093G08 5', mRNA sequence.
ACCESSION BB640911
VERSION BB640911.1 GI:16476069
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 767)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .767
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730093G08"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 175 a 239 c 176 g 172 t 5 others
ORIGIN

Query Match 13.7%; Score 366.6; DB 9; Length 767;
Best Local Similarity 77.6%; Pred. No. 1.4e-58;
Matches 484; Conservative 0; Mismatches 130; Indels 10; Gaps 3;

```
Qy       1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCT------CCCCTGCAGCTCCCCGCG 54
          ||||||| ||||| ||||||| ||  |  | |||          ||| |   ||     |
Db      51 GCGCGCGGGGCGCTGCAGCTCTGGAGCCCTGGATCTTTCCTCACCCGGTGCCTTGAGACC 110

Qy      55 CCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGC 114
           || |||||||| ||||  | | | |  ||||   | || |   ||||  ||||||||||
Db     111 CCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTACGCTACTAGGACAGCGC 170

Qy     115 CGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGG 174
           |||||||  ||||| |  ||  |  | || ||  |||||||||  |||||| || || |
Db     171 CGCCACCACCGCCAG--CCCTCGTCCCGCCCGCGCCGCGCTGCCCTCCCGCGCCCTGGAT 228

Qy     175 ATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAG 234
           ||  |    | || ||| || |||| ||||| |||||||| | || |||||   || |
Db     229 GGAGCTTTCCCTCGCTGGGATCGCTTTANAATGTGATAACTCACGCATCTGATAGCACCG 288

Qy     235 TGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATT 294
           ||  |||||  | ||    |||   ||||| ||||| |||||||||| | |||||||||
Db     289 TGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTTCTCTACACTGAAAAATT 346

Qy     295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354
           ||||||||||||||||||| |||||||||||||| |||| |||||||| |||||||||
Db     347 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 406

Qy     355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414
           ||  |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||
Db     407 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 466

Qy     415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474
           |||||  ||  ||  || |||| |||||||||||||| ||||| |||||  ||||||||||
Db     467 ATGGAGTCTGCAAGTTCTGGGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 526

Qy     475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534
           |||||  |||||||| |||||||| |||||  ||||||||||||||||||||||||||||
Db     527 TTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACA 586

Qy     535 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 594
           |||  ||||||||||||||| |||||  ||  |||||||  |||||  | || ||| || ||
Db     587 TCCGGCCAAGCAATGTTCAATATCCAGGGGGGCCATCTTACAACAAAACGGTCATTGGAC 646

Qy     595 AACAAATCCCATGCAGTTCTGATC 618
           |||||||||| |  |  ||||||
Db     647 AACAAATCCCAGGGGGGCCTGATC 670
```

RESULT 13
AQ418233/c
LOCUS AQ418233 646 bp DNA linear GSS 23-MAR-1999
DEFINITION RPCI-11-165L16.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-165L16, DNA sequence.
ACCESSION AQ418233
VERSION AQ418233.1 GI:4475957
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-165L16.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Homo sapiens"
                     /db_xref="GDB:7563255"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-165L16"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      196 a    116 c    103 g    228 t      3 others
ORIGIN

  Query Match             13.3%;  Score 357;  DB 12;  Length 646;
  Best Local Similarity   99.2%;  Pred. No. 8.4e-57;
  Matches  357;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    2317 GGATATAAGATTTGGCTCATAATGATGAGCCCTATCATTTGATTTGAGTTCTATCATTTA 2376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     646 GGATATAAGATTTGGCTCATAATGATGAGCCCTATCATTTGATTTGAGTTCTATCATTTA 587

Qy    2377 AGAGAGCCTAAATAAAATTATCATCAAGGTATTAAATATAAGACGTTAAATATAATAAAG 2436
           |||||||||||||  |||||||||  |||||||||||||||||||||||||||||  ||
Db     586 AGAGAGCCTAAATANAATTATCATCNAGGTATTAAATATAAGACGTTAAATATAATANAG 527

Qy    2437 TGGGGATATATAGAAAACACACAGTGTTAGCACAGAGTAAGATCTCAATGCACATTTGTT 2496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     526 TGGGGATATATAGAAAACACACAGTGTTAGCACAGAGTAAGATCTCAATGCACATTTGTT 467

Qy    2497 GGATGAATAAATAAATGCAATTGAATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAG 2556
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     466 GGATGAATAAATAAATGCAATTGAATTCCCAGAAAAATGATTGTTTCAAGGAAGTGACAG 407

Qy    2557 TTCTACTTTAGAAGTACTAATTGGAGATGACTTTTATATCCCATTTTGGTAATTATTCAT 2616
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     406 TTCTACTTTAGAAGTACTAATTGGAGATGACTTTTATATCCCATTTTGGTAATTATTCAT 347

Qy    2617 ACATAGCACATATGACCATGATGTTCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 2676
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     346 ACATAGCACATATGACCATGATGTTCAGGGCTTTATAGAACCAAATAAACCTACCATTAC 287


RESULT 14
BB651409
LOCUS       BB651409                 628 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB651409 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230079N02 5', mRNA sequence.
ACCESSION   BB651409
VERSION     BB651409.1  GI:16485662
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C230079N02"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 479.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
```

3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 129 a 200 c 143 g 152 t 4 others
ORIGIN

Query Match 12.2%; Score 325.6; DB 9; Length 628;
Best Local Similarity 76.8%; Pred. No. 6.3e-51;
Matches 437; Conservative 0; Mismatches 122; Indels 10; Gaps 3;

```
Qy       1 GCGCGCGCGGCGCAGCAGCTCCAGATTCACTGCTCT------CCCCTGCAGCTCCCCGCG 54
Db      51 GCGCGCGGGGCGCTGCAGCTCTGGAGCCCTGGATCTTTCCTCACCCGGTGCCTTGAGACC 110

Qy      55 CCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGCGC 114
Db     111 CCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTACGCTACTAGGACAGCGC 170

Qy     115 CGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGG 174
Db     171 CGCCACCACCGCCAG--CCCTCGTCCCGCCCGCGCCGCGCTGCCCTCCCGCGCCCTGGAT 228

Qy     175 ATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAG 234
Db     229 GGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCG 288

Qy     235 TGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATT 294
Db     289 TGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTTCTCTACACTGAAAAATT 346

Qy     295 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 354
Db     347 ATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGC 406

Qy     355 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 414
Db     407 TGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCA 466

Qy     415 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 474
Db     467 ATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTAT 526

Qy     475 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 534
Db     527 GTTAACGAAACTACCATTCTGGNTTAGGTATGGCCATNTGGGCAGACCTNTGACCTTACA 586

Qy     535 TCCTGCCAAGCAATGTTCAACATCCAAGG 563
Db     587 TCCTGCCAAAGCATGTTTCATATCCAAGG 615
```

RESULT 15
CNS059ST/c
LOCUS CNS059ST 1101 bp DNA linear GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 045P22 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL327638
VERSION AL327638.1 GI:8221227
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="045P22"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA045DH11A1~end : T3"
BASE COUNT 287 a 249 c 269 g 278 t 18 others
ORIGIN

Query Match 11.1%; Score 296.8; DB 12; Length 1101;
Best Local Similarity 58.3%; Pred. No. 1.6e-45;
Matches 603; Conservative 0; Mismatches 423; Indels 9; Gaps 5;

```
Qy     344 TTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGA 403
Db    1064 TCATCGTGGGGTGCTNTGTGACTGTGTNNTTTATGTACTNTAATCCATCCACCAGCT-GN 1006

Qy     404 TCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCA 463
Db    1005 TTTCCGGTCCTGTGGAGGCAGGACTGAGCTCGGAGCATGTCAAGAATCAACTATCAACCA 946

Qy     464 AAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCT 523
Db     945 AGAGTGATAAA---AACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCT 889

Qy     524 TTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACC 583
Db     888 TCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACA 829

Qy     584 GTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATC 643
Db     828 GAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACC 769

Qy     644 TGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGG 703
Db     768 TGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGG 709

Qy     704 AATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGA 763
Db     708 AGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGA 649

Qy     764 CTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATC 823
Db     648 ACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATG 589

Qy     824 CCTTCGTG---TTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 880
Db     588 CTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACT 529

Qy     881 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 940
Db     528 GGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCC 469

Qy     941 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 1000
Db     468 ACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCG 409
```

```
Qy   1001 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 1060
          |.   || || || || || || ||||| || || |||||||| || ||| | || ||||
Db    408 CCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAA 349

Qy   1061 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 1120
          | || |||||  |  ||   |  ||  | || || || |||||||| ||| || |||| |
Db    348 AACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGA 289

Qy   1121 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 1180
          |||| |||||    || |  | |||| | || || || |||||||| | |  |||  |||
Db    288 ACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCA 229

Qy   1181 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 1240
            ||||| ||  | ||| |   |    | |||||||  |  |  | |||||||  || |
Db    228 AGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATT 169

Qy   1241 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 1300
          || |   |  ||||    |       ||        |  ||||       |   | |||
Db    168 TTCANGTGGCGAAGACACTTTAAAGCAAGGGCAGTGGCATTTGCCCGAGGC-CACCTGTC 110

Qy   1301 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAA 1359
           |||||| ||| | || |  |||| | |  |||| | ||| |   |  |  ||| | ||
Db    109 AGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAG 50

Qy   1360 TGGTTTTGGAATTAA 1374
          ||||  |||  || |
Db     49 TGGTACTGGGGTTGA 35
```

Search completed: October 6, 2002, 07:16:20
Job time : 2254.54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 54.6005 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-5
Perfect score: 2676
Sequence: 1 gcgcgcgcggcgcagcagct.....ccaaataaacctaccattac 2676

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	4.7	1814	2	US-08-483-151-1	Sequence 1, Appli
2	127	4.7	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	113.4	4.2	1654	1	US-07-914-281-13	Sequence 13, Appl
4	113.4	4.2	1654	1	US-08-393-246-13	Sequence 13, Appl
5	113.4	4.2	1654	1	US-08-525-058A-13	Sequence 13, Appl
6	113.4	4.2	1654	2	US-08-696-731-13	Sequence 13, Appl
7	113.4	4.2	1654	4	US-09-042-531-13	Sequence 13, Appl
8	112.2	4.2	1086	1	US-07-914-281-12	Sequence 12, Appl
9	112.2	4.2	1086	1	US-08-393-246-12	Sequence 12, Appl
10	112.2	4.2	1086	1	US-08-525-058A-12	Sequence 12, Appl
11	112.2	4.2	1086	2	US-08-696-731-12	Sequence 12, Appl
12	112.2	4.2	1086	4	US-09-042-531-12	Sequence 12, Appl
13	112.2	4.2	2042	3	US-09-063-237-2	Sequence 2, Appli
14	112.2	4.2	2043	1	US-07-914-281-1	Sequence 1, Appli
15	112.2	4.2	2043	1	US-08-393-246-1	Sequence 1, Appli
16	112.2	4.2	2043	1	US-08-273-411-4	Sequence 4, Appli
17	112.2	4.2	2043	1	US-08-525-058A-1	Sequence 1, Appli
18	112.2	4.2	2043	2	US-08-696-731-1	Sequence 1, Appli
19	112.2	4.2	2043	4	US-09-042-531-1	Sequence 1, Appli
20	112.2	4.2	2043	5	PCT-US91-00899-1	Sequence 1, Appli
21	111.8	4.2	1316	1	US-07-914-281-10	Sequence 10, Appl
22	111.8	4.2	1316	1	US-08-393-246-10	Sequence 10, Appl
23	111.8	4.2	1316	1	US-08-525-058A-10	Sequence 10, Appl
24	111.8	4.2	1316	2	US-08-696-731-10	Sequence 10, Appl
25	111.8	4.2	1316	4	US-09-042-531-10	Sequence 10, Appl
26	100.8	3.8	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	100.8	3.8	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	100.8	3.8	1488	1	US-07-914-281-9	Sequence 9, Appli
29	100.8	3.8	1488	1	US-08-393-246-9	Sequence 9, Appli
30	100.8	3.8	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	100.8	3.8	1488	2	US-08-696-731-9	Sequence 9, Appli
32	100.8	3.8	1488	4	US-09-042-531-9	Sequence 9, Appli
33	100.8	3.8	2175	4	US-08-482-073-9	Sequence 9, Appli
34	100.8	3.8	2861	4	US-08-482-073-10	Sequence 10, Appl
35	100.8	3.8	3647	1	US-07-914-281-7	Sequence 7, Appli
36	100.8	3.8	3647	1	US-08-393-246-7	Sequence 7, Appli
37	100.8	3.8	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	100.8	3.8	3647	2	US-08-696-731-7	Sequence 7, Appli
39	100.8	3.8	3647	4	US-09-042-531-7	Sequence 7, Appli
40	100.8	3.8	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	97.6	3.6	2134	2	US-08-483-151-3	Sequence 3, Appli
42	97.6	3.6	2134	5	PCT-US96-06427-3	Sequence 3, Appli
43	59.8	2.2	19124	2	US-08-487-826B-13	Sequence 13, Appl
44	54.4	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
45	53.4	2.0	5852	1	US-07-867-106-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 4.7%; Score 127; DB 2; Length 1814;
Best Local Similarity 52.3%; Pred. No. 1.8e-20;
Matches 330; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

```
Qy    563 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 622
          | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy    623 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 682
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    683 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 742
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy    743 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 802
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy    803 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 862
Db    788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy    863 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 922
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy    923 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 982
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy    983 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1042
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy   1043 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1099
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy   1100 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1159
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy   1160 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;    APPLICANT:  The General Hospital Corporation
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             4.7%;  Score 127;  DB 5;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 1.8e-20;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy    563 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 622
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy    623 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 682
Db    608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy    683 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 742
Db    668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy    743 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 802
Db    728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy    803 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 862
Db    788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy    863 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 922
Db    845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy    923 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 982
Db    905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy    983 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1042
Db    965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy   1043 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1099
Db   1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy   1100 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1159
Db   1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy   1160 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 1190
Db   1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 3
US-07-914-281-13
; Sequence 13, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;       ADDRESSEE:  P.C.
;       STREET:  1755 Jefferson Davis Highway, Fourth Floor
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/914,281
;       FILING DATE:  19920720
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1654 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-07-914-281-13

  Query Match             4.2%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 2.5e-17;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
          || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
           |      ||     |  ||| |  ||||  | |||||     ||||      |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
          ||   ||  | ||| ||     |  || || || ||  || | ||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
          |||     ||| || ||||          | | | |       |    | |      | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
          |  |  |   |||      |  ||||| |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
          ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
            |     |     |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
          |||||  |||| || || ||  | |||   || ||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
          |   | ||| | ||  | ||||| || ||||| || |   |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
            ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1252 AAG 1254
            ||
Db   1095 GAG 1097


RESULT 4
US-08-393-246-13
; Sequence 13, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-13

  Query Match             4.2%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 2.5e-17;
```

```
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
           || ||| ||||| |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
             |      ||      |  ||| |  ||||  | |||||      ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
             | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
           |||     ||| || ||||           | | | |     |  | |       | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
             |  |  |   |||      | |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
             |      |      |   ||   ||||    ||  | ||    ||||  ||  |||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
           |   | ||| | ||  | ||||| || ||||| || |    |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
             ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    1252 AAG 1254
             ||
Db    1095 GAG 1097


RESULT 5
US-08-525-058A-13
; Sequence 13, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-13

  Query Match             4.2%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 2.5e-17;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
           || ||| ||||| |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
             |      ||      |  ||| |  ||||  | |||||      ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
             | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
           |||     ||| || ||||           | | | |     |  | |       | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
             |  |  |   |||      | |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
             |      |      |   ||   ||||    ||  | ||    ||||  ||  |||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914
```

```
Qy   1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
          |   | ||| | ||  | ||||| || ||||| || |    |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
            ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1252 AAG 1254
            ||
Db   1095 GAG 1097
```

```
RESULT 6
US-08-696-731-13
; Sequence 13, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-13

  Query Match             4.2%;  Score 113.4;  DB 2;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 2.5e-17;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;
```

```
Qy    490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
          || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
           |       ||      |  ||| |  ||||  | |||||     ||||      |  | |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
          || ||  |  | || ||||||||  |||  |  | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
          ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
          |||     ||| || ||||         | | | |     |  | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
          |  |  |  |||     |  |||||  |||||  |||  ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
          |||||  ||    ||  ||| |    ||           |  |  |  |   |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
          |      |      |   ||  ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy   1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
          |||||  |||| || || ||  | |||   || |||||||| || |||||  |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
          |   | ||| | ||  | ||||| || ||||| || |    |    ||||  || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
            ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1252 AAG 1254
            ||
Db   1095 GAG 1097
```

```
RESULT 7
US-09-042-531-13
; Sequence 13, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
```

```
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/042,531
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1654 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-13

  Query Match             4.2%;  Score 113.4;  DB 4;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 2.5e-17;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     490 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 549
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     550 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 606
             |       ||      |  ||| |  ||||  | |||||      ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     607 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 663
           || ||   |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     664 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 723
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     724 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 783
           ||   ||  | ||| ||    |  || || || ||  || | ||||||  | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     784 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 834
           |||     ||| || ||||           | | | |       |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     835 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 894
           |  |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     895 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 954
           |||||  ||    || |||  |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     955 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 1014
               |    |    |    |  ||  ||||   ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    1015 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1071
           |||||  |||| || || ||  | |||    || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    1072 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1131
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    1132 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1191
           |   | ||| | ||  | ||||| || ||||| || |   |    |||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    1192 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1251
           |||| ||| ||||  | |||||  || |     |||||  || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    1252 AAG 1254
           ||
Db    1095 GAG 1097


RESULT 8
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             4.2%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 4.1e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;
```

```
Qy    565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
          | || |||| |   |   |  | ||    | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
          |   ||  |   | |   |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
          |    ||          |  |  |  |    |||||  |   |               |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
          ||  |  |  |   ||||    ||  | ||    |  || || || ||  ||||  || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
          ||  | |||   || ||||||| || |||||  |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
          || || || || |||||  ||    ||    |||||  ||||  |   | |||  |  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
          |||||  || |||||  ||  |   |    ||||  || ||  ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
          |||||  || |       |||||  || ||||||  ||||  || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 9
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             4.2%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 4.1e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
          | || |||| |   |   |  | ||    | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
          |   ||  |   | |   |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
          |    ||          |  |  |  |    |||||  |   |               |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
          ||  |  |  |   ||||    ||  | ||    |  || || || ||  ||||  || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
          ||  | |||   || ||||||| || |||||  |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
          || || || || |||||  ||    ||    |||||  ||||  |   | |||  |  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
          |||||  || |||||  ||  |   |    ||||  || ||  ||   ||| ||| ||||  |
```

```
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
          ||||| || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 10
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
; INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             4.2%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 4.1e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  |  || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
          | || ||||  |   |   |  | ||    | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
          |   ||  |   |  |   |     |          | |  |  |  |   |||
```

```
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
            |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
          |    ||          |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
          || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 11
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
```

```
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             4.2%;  Score 112.2;  DB 2;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 4.1e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
           | || ||||  |   |   |  | ||     | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
          |   ||  |   |  |    |     |         | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
           |   ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
          || |  |  |   ||||    ||  |'||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
          ||  | |||   || |||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
          || || || ||||| ||     ||    |||||  |||| |   |  | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
          |||||  || ||||| || |    |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 12
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match             4.2%;  Score 112.2;  DB 4;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 4.1e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
           | || ||||  |   |   |  | ||     | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
          |   ||  |   |  |    |     |         | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
           |   ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690
```

```
Qy     973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
           || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db     691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
           ||  | |||    || |||||||| || ||||| |     || ||  |   |||  |  |
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
           || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
           ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
           |||||  || |       ||||| || ||||||  |||| || ||
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 13
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             4.2%;  Score 112.2;  DB 3;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 5.3e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
           |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
           |||||  ||| | || |||||  ||||  || ||    |  ||    |   | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
           |  || || || ||  || | ||||||  | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
           |   ||  |   |  |   |     |              | |  |  |  |   |||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
           |  |||||  |||||  |||  ||||||||  || ||    |||||  ||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
           |    ||           |  |  |  |    ||||| |   |                |
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
           || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
           ||  | |||    || |||||||| || ||||| |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
           || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
           ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
           |||||  || |       ||||| || ||||||  |||| || ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             4.2%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 5.3e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
           |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
           |||||  ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
           |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
           |  ||  |    | |    |          | |  |  |  |   |||
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972
           |    ||          |  |  |  |    |||||  |   |                |
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032
           ||  |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089
           ||  | |||   || ||||||||| || |||||  |     || ||  |   |||  |  |
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149
           || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209
           |||||  || ||||| || |   |    ||||  || || ||   ||| ||| |||| |
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254
           |||||  || |      |||||  || ||||||  |||| || ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 15
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             4.2%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 5.3e-17;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     565 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 624
           |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     625 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 681
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     682 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 741
           |||||  ||| | || |||||  ||||  || ||   |  ||   |  | | | ||
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     742 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 801
           |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     802 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 852
```

```
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     853 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 912

Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     913 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 972

Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     973 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1032

Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    1033 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1089

Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    1090 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1149

Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    1150 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1209

Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1210 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1254

Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:21:51
Job time : 68.6005 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 1686.68 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-5_COPY_29_1374
Perfect score: 1346
Sequence: 1 actgctctcccctgcagctc.....agaaatggttttggaattaa 1346

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1346	100.0	3019	9	AB023021	AB023021 Homo sapi
2	1171.2	87.0	2501	9	HSA238701	AJ238701 Homo sapi
3	1086.4	80.7	80247	9	AL512406	AL512406 Human DNA
4	1009.6	75.0	2156	10	AB049819	AB049819 Rattus no
5	986.8	73.3	2139	10	AB015426	AB015426 Mus muscu
6	958.2	71.2	1128	10	AF345993	AF345993 Rattus no
7	933.6	69.4	1705	10	AF230460	AF230460 Cricetulu
8	789	58.6	1707	5	AB035906	AB035906 Gallus ga
9	676.8	50.3	1080	5	AB035905	AB035905 Xenopus l
10	639	47.5	71396	2	AC100674	AC100674 Mus muscu
11	277.8	20.6	1567	5	AB023627	AB023627 Danio rer
12	234.6	17.4	1832	5	AB023628	AB023628 Danio rer
13	222.6	16.5	1092	5	AB035907	AB035907 Xenopus l
14	220.2	16.4	1264	5	AF288369	AF288369 Gallus ga
15	176.2	13.1	2657	5	GGU73678	U73678 Gallus gall
c 16	169.4	12.6	120514	9	HS84D21	AL034348 Human DNA
17	169.4	12.6	191190	2	AL356303	AL356303 Homo sapi
18	169	12.6	1180	9	BC001879	BC001879 Homo sapi
19	145.2	10.8	1429	10	CGU78737	U78737 Cricetulus
20	145.2	10.8	1462	10	AF090450	AF090450 Cricetulu
21	143.6	10.7	1718	10	AF090449	AF090449 Cricetulu
22	137.2	10.2	1044	5	AB035908	AB035908 Gallus ga
23	132	9.8	2293	4	BTA134FTG	X87810 B.taurus al
24	132	9.8	12607	4	BTA132772	AJ132772 Bos tauru
c 25	128	9.5	71396	2	AC100674	AC100674 Mus muscu
26	127.6	9.5	1158	10	AB039126	AB039126 Mus muscu
27	127	9.4	1487	3	AF016899	AF016899 Schistoso
28	127	9.4	1814	6	AR028686	AR028686 Sequence
29	127	9.4	3594	10	MMU45980	U45980 Mus musculu
c 30	127	9.4	117993	2	AC091284	AC091284 Mus muscu
31	125.6	9.3	1613	4	BTA132776	AJ132776 Bos tauru
32	125.6	9.3	1619	4	BTA132775	AJ132775 Bos tauru
33	125.6	9.3	1644	4	BTA132774	AJ132774 Bos tauru
34	125.6	9.3	1873	4	BTA132773	AJ132773 Bos tauru
35	124.6	9.3	1119	9	AF345881	AF345881 Macaca mu
36	122.8	9.1	1170	10	AB039127	AB039127 Mus muscu
37	122.8	9.1	1170	10	AB039131	AB039131 Mus muscu
38	122.8	9.1	1615	10	MMU33457	U33457 Mus musculu
39	122	9.1	1113	9	AF345884	AF345884 Saimiri s
40	121.2	9.0	1170	10	AB039124	AB039124 Mus muscu
41	121.2	9.0	1170	10	AB039125	AB039125 Mus muscu
42	121.2	9.0	1170	10	AB039128	AB039128 Mus muscu
43	121.2	9.0	1170	10	AB039129	AB039129 Mus muscu
44	121.2	9.0	1170	10	AB039130	AB039130 Mus muscu
45	121.2	9.0	1170	10	AB039132	AB039132 Mus muscu

ALIGNMENTS

RESULT 1
AB023021
LOCUS AB023021 3019 bp mRNA linear PRI 23-JUN-1999
DEFINITION Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX, complete cds.
ACCESSION AB023021
VERSION AB023021.1 GI:5139692
KEYWORDS alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S., Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and Narimatsu,H.
TITLE Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX
JOURNAL FEBS Lett. 453, 237-242 (1999)
REFERENCE 2 (bases 1 to 3019)
AUTHORS Kaneko,M., Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSPENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT       937 a    594 c    541 g    947 t
ORIGIN
```

Query Match 100.0%; Score 1346; DB 9; Length 3019;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 60
Db      29 ACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 88

Qy      61 CCCCAGTCGCGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCG 120
Db      89 CCCCAGTCGCGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCG 148

Qy     121 CCGCGCAGCCCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGT 180
Db     149 CCGCGCAGCCCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGT 208

Qy     181 AATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCA 240
Db     209 AATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCA 268

Qy     241 TGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCAT 300
Db     269 TGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCAT 328

Qy     301 TTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAAC 360
Db     329 TTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAAC 388

Qy     361 CTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAA 420
Db     389 CTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAA 448

Qy     421 ACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGC 480
Db     449 ACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGC 508

Qy     481 CATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCC 540
Db     509 CATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCC 568

Qy     541 ATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAG 600
Db     569 ATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAG 628

Qy     601 ACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGA 660
Db     629 ACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGA 688

Qy     661 TTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGT 720
Db     689 TTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGT 748

Qy     721 TTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGA 780
Db     749 TTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGA 808

Qy     781 CGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGG 840
Db     809 CGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGG 868

Qy     841 TTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAA 900
Db     869 TTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAA 928

Qy     901 GCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGA 960
Db     929 GCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGA 988

Qy     961 TTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATT 1020
Db     989 TTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATT 1048

Qy    1021 ACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGAC 1080
Db    1049 ACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGAC 1108

Qy    1081 CATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATT 1140
Db    1109 CATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATT 1168

Qy    1141 ATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTAT 1200
Db    1169 ATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTAT 1228

Qy    1201 ACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAAT 1260
Db    1229 ACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAAT 1288

Qy    1261 CACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTA 1320
Db    1289 CACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTA 1348

Qy    1321 ATTTAGAGAAATGGTTTTGGAATTAA 1346
Db    1349 ATTTAGAGAAATGGTTTTGGAATTAA 1374
```

```
RESULT 2
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
```

```
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match             87.0%;  Score 1171.2;  DB 9;  Length 2501;
  Best Local Similarity   99.7%;  Pred. No. 2.4e-286;
  Matches 1173;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    171 TTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTAC 230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 TTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTAC 68

Qy    231 GTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATT 290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 GTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATT 128

Qy    291 CTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATT 350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 CTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATT 188

Qy    351 TACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTG 410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 TACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTG 248

Qy    411 AAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTG 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 AAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTG 308

Qy    471 TGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATC 530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 TGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATC 368

Qy    531 CAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATC 590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATC 428

Qy    591 CATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTC 650
          |||||||||||||||| |||||||||| || ||||||||||||||||||||||||||||
Db    429 CATCACCGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAAGCTAGGCCACCCTTC 488

Qy    651 CAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATT 710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATT 548

Qy    711 GAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTAT 770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 GAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTAT 608

Qy    771 GGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTG 830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 GGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTG 668

Qy    831 GTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAG 890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAG 728

Qy    891 CTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGAT 950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 CTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGAT 788

Qy    951 AAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATC 1010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATC 848

Qy   1011 CACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTT 1070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 CACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTT 908

Qy   1071 GTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCAT 1130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 GTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCAT 968

Qy   1131 GTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAAC 1190
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAAC 1028

Qy   1191 AATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGA 1250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 AATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGA 1088

Qy   1251 TTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAG 1310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 TTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAG 1148

Qy   1311 TCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1346
          ||||||||||||||||||||||||||||||||||
Db   1149 TCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1184


RESULT 3
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-504J9 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-504J9 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP3-381A2 is at 80148 in this sequence.
The true right end of clone RP11-77B15 is at 100 in this sequence.

FEATURES Location/Qualifiers

source	1. .80247
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="RP11-504J9"
	/clone_lib="RPCI-11.2"
repeat_region	184. .1422
	/note="L2 repeat: matches 875. .2419 of consensus"
repeat_region	1744. .1834
	/note="MIR repeat: matches 166. .258 of consensus"
repeat_region	2950. .2977
	/note="14 copies 2 mer ca 100% conserved"
repeat_region	3411. .3669
	/note="LTR16C repeat: matches 93. .364 of consensus"
repeat_region	4372. .4704
	/note="MLT1A1 repeat: matches 5. .365 of consensus"
misc_feature	complement(7821. .8406)
	/note="match: GSS: Em:AQ550748"
repeat_region	8126. .8184
	/note="MER58B repeat: matches 272. .332 of consensus"
repeat_region	8185. .8696
	/note="MLT2D repeat: matches 1. .513 of consensus"
repeat_region	8779. .8817
	/note="MLT2D repeat: matches 504. .542 of consensus"
repeat_region	8819. .8974
	/note="MER58A repeat: matches 6. .162 of consensus"
misc_feature	complement(9443. .9891)
	/note="match: GSS: Em:B90633"
misc_feature	complement(9452. .9883)
	/note="match: GSS: Em:AQ119695"
repeat_region	9782. .9894
	/note="MIR repeat: matches 92. .205 of consensus"
repeat_region	10089. .10212
	/note="MIR repeat: matches 135. .262 of consensus"
misc_feature	complement(10093. .10671)
	/note="match: GSS: Em:AQ237061"
repeat_region	10223. .10282
	/note="MIR repeat: matches 203. .262 of consensus"
misc_feature	complement(10315. .10631)
	/note="match: GSS: Em:AQ200259"
misc_feature	10805. .11186
	/note="match: GSS: Em:AQ072353"
repeat_region	11657. .11844
	/note="L2 repeat: matches 2546. .2750 of consensus"
misc_feature	complement(11659. .12084)
	/note="match: GSS: Em:AQ518361"
misc_feature	complement(11674. .11987)
	/note="match: GSS: Em:AQ094678"
misc_feature	12083. .12588
	/note="match: GSS: Em:AQ691235"
repeat_region	12596. .12810
	/note="MIR repeat: matches 23. .248 of consensus"
repeat_region	12920. .12986
	/note="L1MA5 repeat: matches 6228. .6295 of consensus"
repeat_region	12987. .13169
	/note="L1MA5 repeat: matches 5852. .6043 of consensus"
repeat_region	13168. .13280
	/note="L1MA8 repeat: matches 5710. .5832 of consensus"
repeat_region	13851. .17218
	/note="L1PB1 repeat: matches 2787. .6155 of consensus"
repeat_region	18903. .18953
	/note="MLT2CB repeat: matches 439. .501 of consensus"
repeat_region	18984. .19433
	/note="MLT2CB repeat: matches 1. .460 of consensus"
misc_feature	complement(19108. .19690)
	/note="match: GSS: Em:AQ563935"
repeat_region	20683. .20801
	/note="AluJo/FRAM repeat: matches 169. .288 of consensus"
repeat_region	20802. .20862
	/note="AluJ/FLAM repeat: matches 3. .66 of consensus"
repeat_region	20912. .21033
	/note="MIR repeat: matches 114. .237 of consensus"
repeat_region	21047. .21128
	/note="41 copies 2 mer ta 63% conserved"
repeat_region	21676. .21859
	/note="L2 repeat: matches 2130. .2308 of consensus"
repeat_region	21890. .23612
	/note="L1P repeat: matches 841. .2570 of consensus"
repeat_region	23601. .26609
	/note="L1PA5 repeat: matches 3107. .6145 of consensus"
repeat_region	26620. .26789
	/note="MER5A repeat: matches 1. .174 of consensus"
repeat_region	28521. .28627
	/note="MER39 repeat: matches 5. .111 of consensus"
repeat_region	28656. .28837
	/note="MER39b repeat: matches 54. .235 of consensus"
repeat_region	28838. .29142
	/note="AluSc repeat: matches 4. .308 of consensus"
repeat_region	29143. .29494
	/note="MER39b repeat: matches 235. .549 of consensus"
repeat_region	30529. .31344
	/note="L1PA3 repeat: matches 5331. .6146 of consensus"
repeat_region	32234. .32340
	/note="L2 repeat: matches 2627. .2750 of consensus"
repeat_region	32399. .32578
	/note="MIR repeat: matches 12. .203 of consensus"
repeat_region	32816. .32979
	/note="MIR repeat: matches 91. .259 of consensus"
misc_feature	complement(33006. .33450)
	/note="match: GSS: Em:AQ261227"
misc_feature	complement(33058. .33444)
	/note="match: GSS: Em:AQ040453"
repeat_region	33312. .33596
	/note="L1M4 repeat: matches 3473. .3773 of consensus"
repeat_region	34856. .34915
	/note="30 copies 2 mer aa 68% conserved"
misc_feature	complement(34987. .35547)
	/note="match: GSS: Em:AQ321598
	match: STS: Em:G55571"
misc_feature	complement(36488. .36936)
	/note="match: GSS: Em:B41073"
misc_feature	complement(36654. .37090)
	/note="match: GSS: Em:AQ772584"
misc_feature	37102. .37591
	/note="match: GSS: Em:AQ489567"
misc_feature	complement(37581. .38221)
	/note="match: GSS: Em:AQ418233"

```
     repeat_region   37720. .37772
                     /note="L2 repeat: matches 2657. .2710 of consensus"
     misc_feature    complement(37778. .38215)
                     /note="match: GSS: Em:AQ819773"
     misc_feature    38218. .38732
                     /note="match: GSS: Em:AQ599868"
     repeat_region   38861. .39164
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   39612. .39653
                     /note="21 copies 2 mer ac 90% conserved"
     repeat_region   40854. .40909
                     /note="L2 repeat: matches 2655. .2710 of consensus"
     repeat_region   42169. .42270
                     /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   42411. .42548
                     /note="MER5B repeat: matches 21. .181 of consensus"
     repeat_region   42780. .42938
                     /note="MER5B repeat: matches 17. .177 of consensus"
     repeat_region   43229. .43326
                     /note="MIR repeat: matches 40. .148 of consensus"
     repeat_region   45383. .45428
                     /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                     /note="match: STS: Em:G51995
                     match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                     /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
                     /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                     /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                     /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match             80.7%;  Score 1086.4;  DB 9;  Length 80247;
  Best Local Similarity   99.9%;  Pred. No. 2.2e-264;
  Matches 1087;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     259 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35551 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 35610

Qy     319 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35611 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 35670

Qy     379 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35671 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 35730

Qy     439 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35731 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 35790

Qy     499 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35791 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 35850

Qy     559 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35851 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 35910

Qy     619 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35911 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 35970

Qy     679 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35971 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 36030

Qy     739 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36031 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 36090

Qy     799 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36091 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 36150

Qy     859 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36151 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 36210

Qy     919 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 978
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   36211 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTACTT 36270

Qy     979 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36271 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 36330

Qy    1039 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1098
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36331 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 36390

Qy    1099 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36391 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 36450

Qy    1159 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36451 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 36510

Qy    1219 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1278
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36511 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 36570

Qy    1279 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1338
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36571 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 36630

Qy    1339 GGAATTAA 1346
           ||||||||
Db   36631 GGAATTAA 36638


RESULT 4
AB049819
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
```

```
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN

  Query Match           75.0%;  Score 1009.6;  DB 10;  Length 2156;
  Best Local Similarity  86.8%;  Pred. No. 2.4e-245;
  Matches 1147;  Conservative    0;  Mismatches  169;  Indels    6;  Gaps    3;

Qy      25 CGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCGCGCTCTTAGGACAGC 84
           | ||  |||| || |||| ||| | |  || ||    || |   ||||  |||||||||
Db     110 CCCCGGCGCCACTCTCGCCCCCTGGCTACCCACCGATTCCCGCTACGCTGCTAGGACAGC 169

Qy      85 GCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCC 144
           ||||||  | ||||| | ||  | |  || || ||||| | | |||||| || || |
Db     170 GCCGCCGTCACCGCCAGCCCTCGTC--CCGCCCGCGCCTCTCTGCCCTCACGCGCCCTGG 227

Qy     145 GGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATAC 204
           ||  |       || ||| |||||||||||||| |||| ||||| || |||  | ||
Db     228 ATGGAGCTTTCCCTTGCTGGGATCGATTTAGAATGTGATAAGTCAAGTATCTGA--ACAC 285

Qy     205 AGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAA 264
           |||| ||||||  ||||    |||    |||| ||||||||||||||||||    |||||||
Db     286 AGTGGAGTAGTTCAACACTCATCT--CAGCTTTCCATGATATGTTCTCTACCGTGAAAAA 343

Qy     265 TTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGG 324
           ||||||| |||||||||||||| |||||||||||||| |||| ||||||||||||||||
Db     344 TTATGACCTCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATTATCCTGG 403

Qy     325 GCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTC 384
           |||| ||| |||||||||| || || |||||||| || |||||||||| |||||||||
Db     404 GCTGCTTCGTGGCATGTCTACTTATCTACATCAAGCCCACCAACAGCTGGGTCTTCAGTC 463

Qy     385 CAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATT 444
           ||||||| || || |||||||||||||||||||||||| ||||| ||||  ||||||||||
Db     464 CAATGGAGTCTGCAAGCTCTGTGCTGAAAATGAAAAATTTCTTCTCCAGGAAAACTGATT 523

Qy     445 ATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTA 504
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     524 ATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTA 583

Qy     505 CATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGT 564
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||  |||
Db     584 CATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGACCGCTCGTTGT 643

Qy     565 ACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATT 624
           |||||||||||||||| || ||||| || ||| |||||||||| |||||||||| || |
Db     644 ACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAACT 703

Qy     625 TACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAA 684
           |||||||||| || ||||| ||||| |||||||||||||||||||| || || |||||| |
Db     704 TACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTAGAGTCACCCA 763

Qy     685 CTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCC 744
           ||||||| |||||||||||||||||||||||||||||||| |||||||||| ||||| ||||
Db     764 CTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTGTAACTTATCGCC 823

Qy     745 GTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGT 804
           |||||||||||||| |||||||||| |||||||||||||| ||||||| ||||| ||||
Db     824 GTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGCCCCTTTGTGT 883

Qy     805 TTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC 864
           |||||||||||||||||||||||| || |||||||||||||||||||||||||| || ||||
Db     884 TTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGGAATCCCGAGC 943

Qy     865 ATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGC 924
           |||||||  |||||||||||||||  |||||  |||||  ||  |||||||||||  |||||  ||||
Db     944 ATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGGC 1003

Qy     925 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 984
           |||||||||||||||||||  |||||||||||  |||||||  |||||||||  |||||||||
Db    1004 AAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCTACTTGTAAAT 1063

Qy     985 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 1044
           ||||||||||||||||||  |||||  |||||  ||||||||||  ||||||||  |||||||
Db    1064 TTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATG 1123

Qy    1045 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 1104
           | ||| |||||||  || ||||||||||  |||||||||||||||||||||||||||||||
Db    1124 CATTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAACTATGAGAATT 1183

Qy    1105 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 1164
           ||||||||||  |||||||||||||||||||||||||||||||||||  |||||||||  | ||||
Db    1184 ATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGTGAGTTGGCAA 1243

Qy    1165 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 1224
           | |||||||||||  || ||||||||||||||||||||  ||||||||||  ||||||||||
Db    1244 AATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTCAACTGGAGGA 1303

Qy    1225 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 1284
           |||||||  |||||||| ||  |||||  ||||||||||||||||||||  ||||  |||||||
Db    1304 AGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGATC 1363

Qy    1285 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1344
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 ATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGGTTTTGGAATT 1423

Qy    1345 AA 1346
           ||
Db    1424 AA 1425


RESULT 5
AB015426
LOCUS       AB015426                2139 bp    mRNA    linear   ROD 23-JUN-1999
DEFINITION  Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete
            cds.
ACCESSION   AB015426
VERSION     AB015426.1  GI:3702718
KEYWORDS    Fut9; alpha1,3-fucosyltransferase IX.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K.
            and Narimatsu,H.
  TITLE     Expression cloning and characterization of a novel murine alpha1,
            3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15)
            epitope in brain and kidney
  JOURNAL   J. Biol. Chem. 273 (41), 26729-26738 (1998)
  MEDLINE   98434588
REFERENCE   2  (bases 1 to 2139)
  AUTHORS   Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
```

```
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN
```

```
  Query Match             73.3%;  Score 986.8;  DB 10;  Length 2139;
  Best Local Similarity   90.0%;  Pred. No. 1.5e-239;
  Matches 1069;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    1;

Qy     159 CCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATA 218
           | || ||| || |||||||||| |||||||| | || |||||  || ||| ||||||  |
Db       9 CGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGAGTAGTTCA 68

Qy     219 ACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACA 278
             ||    |||   ||||| ||||| |||||||||| | ||||||||||||||||||||
Db      69 GCACTCATCTC--TGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACA 126

Qy     279 TCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCA 338
           |||||||| |||||||||||||| |||| |||||||| |||||||||| |||||||||
Db     127 TCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCA 186

Qy     339 TGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCC 398
           ||||| |||||||| ||||| || |||||||||||| ||||||||||||||||| || ||
Db     187 TGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCA 246

Qy     399 AGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACT 458
           || |||||||||||||||||| ||||| ||||| |||||||||||||||| ||||||
Db     247 AGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACT 306

Qy     459 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 518
           || |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db     307 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 366

Qy     519 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
           |||||||| |||||||| ||||||||||||||| ||||| ||| ||||||||||||||||
Db     367 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426

Qy     579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 638
           || || ||||| || ||  |||||||||| |||||||||| || |||||||||| ||
Db     427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 486

Qy     639 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 698
           |||||||||| |||||||||||||||||||||||| || ||||| |||||||| ||||||
Db     487 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 546

Qy     699 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 758
           |||||||||||||| |||||||| |||||||||| ||||| ||||||||||||||||||
Db     547 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 606

Qy     759 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 818
           |||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db     607 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 666

Qy     819 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 878
           || ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     667 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 726

Qy     879 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 938
           ||||||||| ||||| ||||| || |||||||||| ||||| || |||||||| ||||||
Db     727 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 786

Qy     939 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 998
           || || || ||||||||| ||||||| |||||||| ||||||||||||||||||| |||
Db     787 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 846

Qy     999 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 1058
           ||||| ||||| ||||| |||||||||| |||||||| |||||||| ||| |||||||
Db     847 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 906

Qy    1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
           || |||||||||| ||||| ||||||||||||||||||||||||||||||||||| |||
Db     907 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 966

Qy    1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 1178
           |||||||||||||||||||| ||||||||||||||| ||||||| ||||||||||||
Db     967 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 1026

Qy    1179 GTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTA 1238
           || |||||||||||||||| |||||||||||||||||||||| |||||||| ||||||
Db    1027 GTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTA 1086

Qy    1239 AATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCAT 1298
           || || ||||| |||||||||||||||||||  |||| ||||| ||||| |||||||||
Db    1087 AACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCAT 1146

Qy    1299 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1346
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194
```

```
RESULT 6
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
```

```
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT       324 a    256 c    228 g    320 t
ORIGIN

  Query Match           71.2%;  Score 958.2;  DB 10;  Length 1128;
  Best Local Similarity  91.2%;  Pred. No. 2.3e-232;
  Matches 1017;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     232 TGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTC 291
           ||||| ||||||||||||||||   ||||||||||||||| |||||||||||||||||||| ||||
Db      14 TGCTTTCCATGATATGTTCTCTACCGTGAAAAATTATGACCTCAACATCCAAAGGCATTC 73

Qy     292 TTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTT 351
           |||||||||||| |||| ||||||||||||||||||||| ||| |||||||||| || || |
Db      74 TTCGCCCATTTCTAATCGTCTGCATTATCCTGGGCTGCTTCGTGGCATGTCTACTTATCT 133

Qy     352 ACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGA 411
           ||||||| || |||||||||||| |||||||||||||||| || || ||||||||||||||
Db     134 ACATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCAAGCTCTGTGCTGA 193

Qy     412 AAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGT 471
           |||||||||| ||||| ||||  |||||||||||||||||||||||||||||||||||| |
Db     194 AAATGAAAAATTTCTTCTCCAGGAAAACTGATTATTTTAATGAAACTACTATTCTGGTTT 253

Qy     472 GGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCC 531
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     254 GGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCCATGTTCAACATCC 313

Qy     532 AAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCC 591
           |||| |||||||||||||||||| ||  |||||||||||||||||| || |||||| |
Db     314 AAGGGTGCCATCTCACAACGGACCGCTCGTTGTACAACAAATCCCATGCGGTCCTGATAC 373

Qy     592 ATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCC 651
           | ||| |||||||||| |||||||||| || |||||||||| || ||||| |||||| |
Db     374 ACCACAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTC 433

Qy     652 AGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTG 711
           |||||||||||||||||| || || ||||| |||||||| ||||||||||||||||||| |
Db     434 AGAAATGGATTTGGATGAACTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTG 493

Qy     712 AGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATG 771
           |||||||||| |||||||||| ||||| |||||||||||||||| |||||||||||| |
Db     494 AGCACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATACAAGTGCCTTACG 553

Qy     772 GCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGG 831
           |||||||||||||| |||||||  ||||| |||||||||||||||||||||||||||| || |
Db     554 GCTTCTTGACGGTGAGCACAAGCCCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAG 613

Qy     832 TGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGC 891
           |||||||||||||||||||||| || |||||||||| ||||||||||||||| ||||
Db     614 TGTGCTGGGTTGTGAGTAACTGGAATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGC 673

Qy     892 TAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATA 951
           | ||||| || |||||||||| ||||| |||||||||||||||||||||||| |||||||
Db     674 TCAGCAAGAGTATTGAAATCCACACCTATGGGCAAGCATTTGGAGAATATGTGAATGATA 733

Qy     952 AAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCC 1011
           ||||| ||||||| |||||||| ||||||||||||||||||||||||||||||| ||||| |
Db     734 AAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTC 793

Qy    1012 ACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1071
           |||| |||||||||| |||||||| |||||||| ||| ||||||| ||||||||||||||
Db     794 ACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGTTCTGTACCTGTTG 853

Qy    1072 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1131
           | ||||| ||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     854 TCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATG 913

Qy    1132 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACA 1191
           ||||||||| |||||||||||||||| | ||||| |||||||||| || ||||||||||
Db     914 TGGAAGATTTTAACTCTCCCAGTGAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACA 973

Qy    1192 ATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGAT 1251
           |||||||||| |||||||||| |||||||||||||||||| |||||||| || ||||| |
Db     974 ATAAGTTATATCTTAGTTACTTCAACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGT 1033

Qy    1252 TTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGT 1311
           ||||||||||||||||||  |||| |||||||||||| ||||||||||||||||||||||
Db    1034 TTTGGGAATCACATGCATGCCTGGCATGCGATCATGTAAAAAGGCATCAAGAATATAAGT 1093

Qy    1312 CTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1346
           ||||||| |||||||||||||||||||||||||||
Db    1094 CTGTTGGCAATTTAGAGAAATGGTTTTGGAATTAA 1128


RESULT 7
AF230460
LOCUS       AF230460                1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
  PUBMED    10700388
REFERENCE   2  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT       515 a    336 c    320 g    534 t
ORIGIN

  Query Match           69.4%;  Score 933.6;  DB 10;  Length 1705;
  Best Local Similarity  89.7%;  Pred. No. 4.5e-226;
  Matches 1026;  Conservative    0;  Mismatches  114;  Indels    4;  Gaps    2;

Qy     203 ACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAA 262
```

```
Db      2 ACAGTGAAGTGGTTCTACACCCGTCT--CTGCTGCCCATG--CTGTTCTCTACAGTAAAA 57

Qy    263 AATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCT 322
Db     58 AATTATGACATCAACATCCAAAGGCATTCTTCGTCCATTCCTAATCGTCTGTGTTATCCT 117

Qy    323 GGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAG 382
Db    118 GGCCTGCTTCATGGCATGTCTGCTTATTTACATCAAGCCCACCAACAGCTGGGTGTTCAG 177

Qy    383 TCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGA 442
Db    178 TCCAATGGAGTCTGCCAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGA 237

Qy    443 TTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCT 502
Db    238 TTATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAGACCTTTGACCT 297

Qy    503 TACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACT 562
Db    298 TACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACAGACCGCTCACT 357

Qy    563 GTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAA 622
Db    358 GTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGATCTGACTAA 417

Qy    623 TTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACC 682
Db    418 CTTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAACCTAGAGTCTCC 477

Qy    683 AACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCG 742
Db    478 CACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACTCTAACATATCG 537

Qy    743 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGT 802
Db    538 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGT 597

Qy    803 GTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGA 862
Db    598 GTTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAACTGGAATCCGGA 657

Qy    863 GCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGG 922
Db    658 GCATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATCCATACCTATGG 717

Qy    923 GCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAA 982
Db    718 GCAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATATCTACTTGTAA 777

Qy    983 ATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAA 1042
Db    778 ATTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAAAAGCTCTACAA 837

Qy   1043 TGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAA 1102
Db    838 TGCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAAAACTATGAGAA 897

Qy   1103 TTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
Db    898 TTATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGTTGGC 957

Qy   1163 AAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAG 1222
Db    958 AAAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTATTTCAACTGGAG 1017

Qy   1223 GAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGA 1282
Db   1018 GAAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGCCTGGCTTGTGA 1077

Qy   1283 TCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAA 1342
Db   1078 TCACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAATGGTTTTGGAA 1137

Qy   1343 TTAA 1346
Db   1138 TTAA 1141
```

```
RESULT 8
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPPFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match           58.6%;  Score 789;  DB 5;  Length 1707;
  Best Local Similarity   82.8%;  Pred. No. 2.3e-189;
  Matches  900;  Conservative    1;  Mismatches  186;  Indels    0;  Gaps    0;

Qy    260 AAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTAT 319
Db    261 ACAAATTATGACATCAACATCTAAAGGAATTTTCCGGCCATTTTTTATTGTCTTCATTGT 320

Qy    320 CCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTT 379
Db    321 YCTTGGTTGTTTCATGGCATTTATACTAATTTATATTAAACCAACAAGCAGCTGGATCTC 380

Qy    380 CAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAAC 439
Db    381 TAATCCTGTGGAATCAGCCAGCTCAGTTTTGAAAATGAAGAGCTTCTTTTCTTCCAAAAC 440

Qy    440 TGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGA 499
Db    441 TGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGACATTCGA 500
```

```
Qy    500 CCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTC 559
          || |||||||||||| | |||||||||||||| || |||||||| || |  ||||| ||
Db    501 TCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTGACCGCTC 560

Qy    560 ACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGAC 619
          ||| || ||||| |||||||||||||| || |||||| | |||||  |  ||||||||||
Db    561 ACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGGATCTGAC 620

Qy    620 AAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATC 679
          |||||||||||||||| |||||||| ||||||||||||||||||||| ||||||||
Db    621 TAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACTTGGAATC 680

Qy    680 ACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTA 739
          |||||||| ||||| |||||||||||||||| ||| | |||||||| || ||||||||
Db    681 TCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCCTGACTTA 740

Qy    740 CCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTT 799
          ||| |||||||||||||| || |||||||||||| | |||  ||  | |||    ||||
Db    741 CCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAGGCGCCTT 800

Qy    800 CGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCC 859
          | |||||||||||| |||| || ||||| || |||||||| ||||||||||||||||
Db    801 TACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACTGGAACCC 860

Qy    860 TGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTA 919
          |||||||| |||||||||||||||||||||| |||||  ||||||||||||||||||
Db    861 TGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCCATACCTA 920

Qy    920 CGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTG 979
          |||||||| |||||||| || ||||| |||||||| || ||||| || || ||  |||
Db    921 CGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCTCCAGTTG 980

Qy    980 TAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATA 1039
          ||||| || ||||| |||||||||||||||||||||| |||||||| || || || || ||
Db    981 CAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGAAACTTTA 1040

Qy   1040 CAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGA 1099
          ||||||| |||||||||| || ||||| ||||| ||||| || || || |||||||| ||
Db   1041 CAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAAACTACGA 1100

Qy   1100 GAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCT 1159
          |||||| |||||||||| || || || || |||||||||| |  || || || |||||
Db   1101 GAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAAGAGAGCT 1160

Qy   1160 AGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 1219
          ||| | |||||    |    | |||||||| || ||| | ||||||||||||| |||||
Db   1161 TGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACTTCAACTG 1220

Qy   1220 GAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTG 1279
          |||||||||||||| | ||  |||||||   ||| |||||||| |||||||| | |||||
Db   1221 GAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTCTTGCTTG 1280

Qy   1280 CGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTG 1339
          |||||||||||||| || || ||||| |||||  |||| |||||||| ||||||||||
Db   1281 TGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAATGGTTTTG 1340

Qy   1340 GAATTAA 1346
          ||| |||
Db   1341 GAACTAA 1347


RESULT 9
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN

  Query Match           50.3%;  Score 676.8;  DB 5;  Length 1080;
  Best Local Similarity  76.7%;  Pred. No. 6e-161;
  Matches  828;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy    267 ATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGC 326
          |||||||| |||| |||||||| | || |||||||| |||| ||  | |  ||  |
Db      1 ATGACATCAGCATCTAAAGGAATCTTACGGGCATTTTTAGCTGTCAGCGGTGTGTTGATC 60

Qy    327 TGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCA 386
          ||||| |||  |||| | || || ||| ||||||| || |||| |||||| | ||||||
Db     61 TGTTTTATGATATGTTTACTGATATACGTCAAACCAACAAACAACTGGATTTCCAGTCCT 120

Qy    387 ATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTAT 446
          || ||||||||    || || || |||||||| || || || ||| | |||    |||||
Db    121 ATCGAATCAGCACATTCAGTCCTTAAAATGAAGAATTTTTTCTCCTCTAAACAGAATTAT 180

Qy    447 TTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACA 506
          |  |||||||     ||||  | ||||||||| ||||||||||| ||||| || | |
Db    181 TACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGACTTTTGAACTGAAA 240

Qy    507 TCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTAC 566
          |||||||  || |||| || ||  | || ||||||||| || || |||||  |||| |||
Db    241 TCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTGACCGCGCACTCTAC 300

Qy    567 AACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTA 626
          ||||| || ||||| || ||  | || ||| ||||||| || |||||||| |  ||  |
Db    301 AACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGGATCTCATTAACCTG 360

Qy    627 CCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACT 686
          || || ||    |||||||| |||||||||||||||||||||||| ||||||||||||
Db    361 CCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATCTGGAATCACCAACA 420

Qy    687 CACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGT 746
          || || || ||||| |||||||||||||||| |||||||||||||||||||| | |||
Db    421 CATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTCTGACTTACAGGCGT 480

Qy    747 GATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
          || |||||||||||||||||||||||||||||| ||   |||||||||| || || |  |||
```

```
Db    481 GACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAAAACCATTTGACTTT 540

Qy    807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
          || |||||||||||||| || || || || |||||||| || |||||||||| || |||
Db    541 GAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACTGGAACCCAGAACAT 600

Qy    867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
          |||||||||| |||||||||| || | ||||  ||||||||    || || ||||||
Db    601 GCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTATAACTTATGGGCAA 660

Qy    927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
          || ||||| |||||  | |  |||||||   || |||| || ||  |  |||| ||||||
Db    661 GCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTGCATCTTGCAAATTT 720

Qy    987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 1046
          ||||| || |||||||||| || ||||| ||||| ||||| ||||| || || |||||
Db    721 TATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAAAACTTTATAATGCA 780

Qy   1047 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1106
          ||  ||||||| || || ||  |||| |||||||  |||| ||||||||||||||| ||
Db    781 TTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAAACTATGAGAACTAC 840

Qy   1107 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1166
          ||||| |||||||| || || || |||||||||| |  |||||| || ||  ||||  |
Db    841 ATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAAGAGAATTAGCTGAC 900

Qy   1167 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
           ||||        |  ||||| | |   |    ||||| || ||||  ||||||||||
Db    901 CATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACTACAACTGGAGGAAA 960

Qy   1227 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1286
           |||| || || ||| | || |  || |||||||| |||||||| || || || || |||
Db    961 CATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCTTAGCGTGTGACCAT 1020

Qy   1287 GTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1346
          |||||||||||||||||| || || || || ||| ||||||| ||||| ||||| |||
Db   1021 GTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGTGGTTCTGGAACTAA 1080
```

```
RESULT 10
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1702: contig of 814 bp in length
            *     1703 1802: gap of      100 bp
            *     1803     2642: contig of 840 bp in length
            *     2643 2742: gap of      100 bp
            *     2743     3637: contig of 895 bp in length
            *     3638 3737: gap of      100 bp
            *     3738     4635: contig of 898 bp in length
            *     4636 4735: gap of      100 bp
            *     4736     5657: contig of 922 bp in length
            *     5658 5757: gap of      100 bp
            *     5758     6692: contig of 935 bp in length
            *     6693 6792: gap of      100 bp
            *     6793     7703: contig of 911 bp in length
            *     7704 7803: gap of      100 bp
            *     7804     8742: contig of 939 bp in length
            *     8743 8842: gap of      100 bp
            *     8843     9787: contig of 945 bp in length
            *     9788 9887: gap of      100 bp
            *     9888    10819: contig of 932 bp in length
            *    10820 10919: gap of      100 bp
            *    10920    11808: contig of 889 bp in length
            *    11809 11908: gap of      100 bp
            *    11909    12830: contig of 922 bp in length
            *    12831 12930: gap of      100 bp
            *    12931    13847: contig of 917 bp in length
            *    13848 13947: gap of      100 bp
            *    13948    14886: contig of 939 bp in length
            *    14887 14986: gap of      100 bp
            *    14987    15911: contig of 925 bp in length
            *    15912 16011: gap of      100 bp
            *    16012    16903: contig of 892 bp in length
            *    16904 17003: gap of      100 bp
            *    17004    17906: contig of 903 bp in length
            *    17907 18006: gap of      100 bp
            *    18007    18931: contig of 925 bp in length
            *    18932 19031: gap of      100 bp
            *    19032    19945: contig of 914 bp in length
            *    19946 20045: gap of      100 bp
            *    20046    20951: contig of 906 bp in length
            *    20952 21051: gap of      100 bp
            *    21052    21949: contig of 898 bp in length
```

```
*    21950 22049: gap of      100 bp
*    22050    22950: contig of 901 bp in length
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp

 Query Match             47.5%;  Score 639;  DB 2;  Length 71396;
 Best Local Similarity   85.0%;  Pred. No. 8e-151;
 Matches  764;  Conservative    0;  Mismatches  122;  Indels   13;  Gaps    4;

Qy     338 ATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGC 397
           || |||   | | |      |  || ||||||||||| | |||||||||||||||| || ||
Db   65456 ATCTCTCGCCCTGTGGTGGATTCCCACCAACAGCTCGGTCTTCAGTCCAATGGAGTCTGC 65515

Qy     398 CAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAAC 457
           || |||||||||||||||||||| ||||| ||||| |||||||||||||||||| |||||
Db   65516 AAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAAC 65575

Qy     458 TACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 517
           ||| |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db   65576 TACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGC 65635

Qy     518 AATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCA 577
           |||||||||| |||||||| |||||||||||||||| ||||| ||| ||||||||||||||||
Db   65636 AATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCA 65695

Qy     578 TGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGC 637
           ||| || ||||| || ||  |||||||||| |||||||||| || |||||||||| ||
Db   65696 TGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGC 65755

Qy     638 TAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCA 697
           |||||||||| |||||||||||||||||||||||| || ||||| |||||||| |||||
Db   65756 CAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCA 65815

Qy     698 AAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATAT 757
           ||||||||||||||| |||||||| |||||||||| ||||| ||||||||||||||||||
Db   65816 AAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATAT 65875

Qy     758 CCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAG 817
           |||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||
Db   65876 CCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAG 65935

Qy     818 CAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAA 877
           ||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65936 CAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAA 65995

Qy     878 GTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGA 937
           ||||||||| ||||| ||||| || |||||||||| ||||| || |||||||| |||||
Db   65996 GTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGA 66055

Qy     938 ATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTT 997
           ||| || || |||||||||| ||||||| |||||||||| ||||||||||||||||||  |
Db   66056 ATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCTCT 66115

Qy     998 TGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGG 1057
             ||| |      ||  |  | | |   | |         | ||     ||| |||||||
```

```
Db  66116 GAAAACTCATTCACAAGATTACTCCCAAAAG---------TCCACATGCTTTTTGGCTGG 66166

Qy   1058 CTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGA 1117
          || ||||||||||| ||||| ||||||||||||||||||||||||||||||||||||| ||
Db  66167 TTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGA 66226

Qy   1118 TTCATTCATTC--ATGTGGAAGATTATAACTCT-CCCAGTGAGCTAGCAAAGTATCTGAA 1174
          ||||||||||    |||||||||| | |||| |||||||| |||||||    |||||
Db  66227 TTCATTCATTCCATCGTGGAAGATTTTCACTCTCCCCAGTGAGTTAGCAAATATCCTGAA 66286

Qy   1175 GGAAGTCGACAAAAACAATAAGTT-ATACCTTAGTTACTTTAACTGGAGGAAGGATTTC 1232
          |||||||||||| ||||||||||  |||||||||||| ||||| |  ||| |  |||
Db  66287 GGAAGTCGACAANAACAATAAGTTGGTACCTTAGTTACTCTAACTCGNAGAAAGGATTC 66345


RESULT 11
AB023627
LOCUS       AB023627               1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN

  Query Match           20.6%;  Score 277.8;  DB 5;  Length 1567;
  Best Local Similarity   58.5%;  Pred. No. 1.3e-59;
  Matches  527;  Conservative    0;  Mismatches  362;  Indels   12;  Gaps    2;

Qy    449 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 508
          | | ||||| |   |  || | |||||||  || || ||       ||||||  |   |
Db    461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy    509 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 568
            ||     |  ||||||||||||| || || ||| | |||   ||| |     ||  | |

Db    521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy    569 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 628
          ||| | |||| |||| || | || ||| |||| | ||   ||         |||| ||
Db    581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGG---------ATTTGCC 631

Qy    629 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 688
          || |   || | ||| | ||||| || ||||| |||  |||| |||| || || | | |
Db    632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy    689 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 748
            | |  | ||| ||    ||  |  |   |||||| |||||    | || ||||  ||
Db    692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy    749 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTT 805
          |||||||| | |||||||||    | |  || |     |  | |     | |        ||
Db    752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy    806 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 865
            | || |  || || ||| | || |||||| | || |||||||    |      | ||
Db    812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy    866 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 925
              |    |||| || |||||| |    |||  |||| || |  |  |||| ||
Db    872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy    926 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 985
            ||      |   || |||||| |  |||      |    ||||   |||||||||
Db    932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy    986 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 1045
          || |||||||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db    992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy   1046 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 1105
          |||   |   ||  | || |||||||| || || || || || || |||||||||  | |
Db   1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy   1106 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 1165
          ||| |||  |||  | ||||| |||||||| ||||      ||||||  || || ||  |
Db   1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy   1166 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 1225
            | ||  |     | ||| ||||| |  |    |||       || || |||||||| ||
Db   1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy   1226 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 1285
            |||| | |  ||    | ||  |||| ||      ||||||  |||  | |||| ||| |
Db   1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy   1286 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1345
          | | |||||| || |  | ||||   || ||     || |  |    |||||  |||  |||
Db   1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy   1346 A 1346
          |
Db   1352 A 1352


RESULT 12
AB023628
LOCUS       AB023628               1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
```

```
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match             17.4%;  Score 234.6;  DB 5;  Length 1832;
  Best Local Similarity   55.5%;  Pred. No. 1.2e-48;
  Matches  500;  Conservative    0;  Mismatches  389;  Indels   12;  Gaps    2;

Qy     449 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 508
           | | || |   |    | || | ||| ||   || ||  |      | ||||||| | |
Db     619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy     509 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 568
             ||          ||  |||||||| || || ||   | |||    ||| |    |||  | |
Db     679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy     569 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 628
            ||| | |||| |||| || | || ||| |||| || ||           ||| |||| ||
Db     739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGC---------ACAGATTTTCC 789

Qy     629 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 688
            || |   |  | |||   ||||| || ||||| |||  |||| |||| || || || ||
Db     790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy     689 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 748
               |||||      | |  |  |   |||||| ||||    | |||| | | ||
Db     850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy     749 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA---AATCCCTTCGTGTT 805
           ||||||||| | |||||||| ||      |    |    ||||    | |        |||
Db     910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy     806 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 865
             | ||  |  || || ||| | || ||||||  | || |||||||    |     | ||
Db     970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy     866 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 925
               |     | |  | | | ||| |    |||  |||| |  |    |||||  ||
Db    1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy     926 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 985
               ||     |     | |||| | |  |||       |  | ||||   |||||||||
```

```
Db    1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy     986 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 1045
           || || |||||||| || || || |||| || ||  |||| || ||||| | |||| |
Db    1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy     1046 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 1105
            |||        ||  | || ||||| ||  | || ||  | ||  | |||||||  | |
Db     1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy     1106 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 1165
             || |||  |||  | ||||| |||||||| ||||      ||||||  || || ||  |
Db     1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy     1166 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 1225
              | || |       | ||| ||||| |  |    |||       |  || ||||||||||
Db     1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy     1226 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 1285
              ||     |  |||   | ||   ||| ||     |||| | |||  |  ||| ||| |
Db     1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy     1286 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 1345
            | | |||||| || |  | |||   | |||    || |  |    ||||| |||  |||
Db     1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy     1346 A 1346
              |
Db     1510 A 1510
```

```
RESULT 13
AB035907
LOCUS       AB035907                1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
```

```
                    EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN

  Query Match           16.5%;  Score 222.6;  DB 5;  Length 1092;
  Best Local Similarity  57.3%;  Pred. No. 1.2e-45;
  Matches  484;  Conservative    0;  Mismatches  349;  Indels   12;  Gaps    4;

Qy    450 AATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 509
          ||  ||   ||| |  |  | |||     |||||||||  |           |||
Db    199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy    510 TGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAAC 569
          |||||     | |||||  ||   || |||||  |||||||  || |    || ||| |
Db    259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy    570 AAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCT 629
            |  | | ||  || || |||| ||||| |||||   ||  ||| | |   |  | |||
Db    319 GAGGCAGACGCCATTATGTTCCACCACCGGGACAT---TTCTGATTTCAGTGACCTTCCT 375

Qy    630 CAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 689
            | |    | |    || ||| || ||||| |||||||| || ||||| ||  | ||
Db    376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy    690 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 749
          | |||     |    |  | | |  | | || |||  |||  || | ||| |    ||
Db    436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy    750 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 809
          || ||||||    |||| || ||||||||     |   ||  || |    |    | |
Db    496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy    810 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 869
          ||||  ||||    ||  |||||  ||||||  | || |||||||||   || || |
Db    553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy    870 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 929
          || ||| |||||||||| ||||| || ||   | | |||||| |    ||||| |
Db    613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy    930 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 989
          | |||     || | || ||  | ||  |  |    ||| | || |  | |||||||||
Db    670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy    990 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 1046
          |  | ||||| |||||  | |||| ||||||||||||||||||||| |     || ||
Db    730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy   1047 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1106
          |||    |  | |  | ||  | ||  |||| ||      |  | ||||| |||| |
Db    790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy   1107 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1166
          || ||       || |||||||| ||||| ||||    |   |||||  |||| || | |
Db    850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy   1167 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          || |||||| |  | || |||||| | | || ||||  || || || |  |||| ||||
Db    910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy   1227 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1286
            || |   ||  |  |  |    || |||||    |||   ||    || ||| |
Db    970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy   1287 GTGAA 1291
          || ||
Db   1030 GTCAA 1034


RESULT 14
AF288369
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.

ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN

  Query Match           16.4%;  Score 220.2;  DB 5;  Length 1264;
  Best Local Similarity  58.0%;  Pred. No. 5e-45;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    6;  Gaps    3;

Qy    459 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 518
          || || ||| ||||   ||||| || || ||   |||| |  | |    ||||  ||
Db    175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy    519 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
            | ||  ||| ||     |||||  |||| ||  ||||      || | |  ||  |  ||
Db    235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy    579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 638
          || ||  | || ||||||  | ||  |   ||||||        |   |  | | |    |
Db    295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy    639 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 698
          ||||  || | ||||   |||||  ||| | ||  |||| || |||| |||  ||||  |
Db    355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy    699 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 758
            ||| ||| || ||| | || |||||||||| || | |||||  | || || || |||
Db    415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy    759 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 818
              || |||||      ||  | |   ||   | |||       ||      | |||  |
Db    475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy    819 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 878
          ||    ||  ||||  ||||||  || |||||  ||||  ||          | | || |||
Db    535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy    879 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 938
          || ||| | ||||| |  ||| |||      |  ||   |||||||  |||      |
Db    595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGC-GGCAACACTTGCC 653
```

```
Qy     939 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 998
           |||| |||   | |   |     |  ||| ||||    |  || || ||  |  |||||
Db     654 CCTGTCCATG--GACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy     999 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATAC---AATGCTTTTCTGGCT 1055
           || || |||   ||||| || |||||||| || ||||| |     |||||  |     ||
Db     712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy    1056 GGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCA 1115
           ||| |  | || ||||||||||| ||| || | |||||||||||| | |   | ||  |
Db     772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy    1116 GATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAG 1175
           || || ||||| || || || || | |  | | | |||| || ||  |||| ||| ||
Db     832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy    1176 GAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
           ||| |   ||   |||  ||  |||| | |||||  | ||| | | |
Db     892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942


RESULT 15
GGU73678
LOCUS       GGU73678                2657 bp    DNA     linear   VRT 20-DEC-1996
DEFINITION  Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete
            cds.
ACCESSION   U73678
VERSION     U73678.1  GI:1657998
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N.,
            Schultz,T.L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
  TITLE     Molecular cloning and characterization of CFT1, a developmentally
            regulated avian alpha(1,3)-fucosyltransferase gene
  JOURNAL   J. Biol. Chem. 271 (51), 32960-32967 (1996)
  MEDLINE   97115837
REFERENCE   2  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M. and Larsen,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical
            Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES             Location/Qualifiers
     source          1. .2657
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            170. .1240
                     /gene="CFT1"
     CDS             170. .1240
                     /gene="CFT1"
                     /note="similar to mammalian fucosyltransferase FucTIV"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase"
                     /protein_id="AAC60060.1"
                     /db_xref="GI:1657999"
                     /translation="MELGPRWSPAARPGCPRRWRRRWALLGALLGAALALYVCVRELR
                     RRGSAAGRPEGEVTVLLWWEPFGRPWRPADCRRRYNITGCLLSADRGRYGEARAVLFH
                     HRDLALHGRQGLPRGPPPRPPRQRWVWMNFESPSHSPGLRGLAGLFNWTMSYRRDSDV
                     FVPYGYLYEPPSPRPFVLPRKSRLVAWVISNWNEEHARVRYYRQLKEHLPIDVYGARG
                     MALLEGSVVKTVSAYKFYLAFENSQHTDYITEKLWKNAFAASAVPVVLGPRRANYERF
                     IPADSFIHVDDFPSPRLLATYLKFLDKNKPSYRRYFAWRNKYEVHVTSFWDEHYCKVC
                     EAVRTAGNQLKTVQNLAGWFES"
BASE COUNT      581 a    712 c    818 g    546 t
ORIGIN

  Query Match           13.1%;  Score 176.2;  DB 5;  Length 2657;
  Best Local Similarity  53.1%;  Pred. No. 9.2e-34;
  Matches  451;  Conservative    0;  Mismatches  383;  Indels   15;  Gaps    3;

Qy     459 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 518
           ||  | ||| ||||  |  ||| || || | |||    ||    |  |||||
Db     341 ACGGTGCTGCTGTGGTGGGAGCCCTTCGGCCGCCCCTGGCGCCCGGCCGACTGCCGGAGG 400

Qy     519 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
             | |||||||   || ||||  || |   | |||||    | ||||  | |  | |
Db     401 CGCTACAACATCACGGGCTGCCTCCTGAGCGCCGACCGCGGGCGGTACGGCGAGGCTCGG 460

Qy     579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAG-- 636
           || || ||  |||| ||||| ||| |      | |          |   |  |  |
Db     461 GCCGTGCTCTTCCACCACCGCGACCTGGCGCTGCACGGCCGCCAGGGGCTGCCCCGCGGG 520

Qy     637 -------CTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 689
                 |  | || ||    |||  ||| | |||||||| || || || || | |||
Db     521 CCCCCGCCGCGACCCCCGCGGCAGCGGTGGGTGTGGATGAACTTCGAGTCGCCCTCGCAC 580

Qy     690 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 749
           | |||     | | ||| | |       | || |||  |||  || | ||||| || ||
Db     581 TCCCCCGGGCTGCGGGGCCTGGCCGGGCTCTTCAACTGGACCATGTCGTACCGACGGGAC 640

Qy     750 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 809
           || ||  ||   ||||| || || | | |    |       |    ||       || |
Db     641 TCGGACGTCTTCGTGCCCTACGGGTACCTCTACGAGCCGCCGTCGCCGCGGCCCTTCGTG 700

Qy     810 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 869
           ||||  |||||       |||||  ||||||  | || |||||||||   ||||| |||
Db     701 CTGCCTCGCAAAAGCCGGCTGGTGGCCTGGGTCATCAGCAACTGGAACGAGGAGCACGCC 760

Qy     870 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 929
           | ||    || |||    |||| |   |  | |    ||| |    ||||||   ||
Db     761 CGCGTGCGCTACTACCGGCAGCTGAAGGAGCACCTGCCCATCGACGTGTACGGG---GCG 817

Qy     930 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 989
           ||          |   ||    |  || |    ||  | || || |  || || ||
Db     818 CGCGGGATGGCGCTGCTGGAGGGCAGCGTGGTGAAGACGGTGTCGGCCTACAAGTTCTAC 877

Qy     990 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 1046
           ||  |||| || || ||    |||| ||| |||||||| || ||||| |   | || ||
Db     878 CTGGCCTTCGAGAACTCCCAGCACACGGACTACATCACCGAGAAGCTGTGGAAGAACGCC 937

Qy    1047 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 1106
           ||    ||  || | || |||||||| || || ||    ||||  ||||| |||   |
Db     938 TTCGCCGCCAGCGCGGTGCCTGTTGTCCTCGGCCCCCGCAGGGCCAACTACGAGCGCTTC 997

Qy    1107 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 1166
           || || || || || ||||| || ||||| || |   |  ||  |  ||| || |
Db     998 ATCCCCGCCGACTCCTTCATCCACGTGGACGACTTCCCCAGCCCACGGCTGCTGGCCACC 1057

Qy    1167 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
           || ||||||    |||| ||||||||        |||   || || ||   |||| ||||
Db    1058 TACCTGAAGTTCCTCGATAAAAACAAGCCCAGCTACAGGAGGTATTTCGCCTGGCGGAAC 1117

Qy    1227 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 1286
           | |     || |  |  |    || |||||    |||  ||   || ||||||
Db    1118 AAGTATGAAGTCCACGTCACGTCTTTCTGGGATGAGCATTACTGCAAGGTTTGCGAGGCC 1177

Qy    1287 GTGAAAAGG 1295
           ||||  | |
Db    1178 GTGAGGACG 1186


Search completed: October  6, 2002, 04:46:41
Job time : 1826.68 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 151.049 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-5_COPY_29_1374
Perfect score: 1346
Sequence: 1 actgctctcccctgcagctc.....agaaatggttttggaattaa 1346

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1346	100.0	2676	21	AAZ92647	Human alpha-1,3-fu
2	1088	80.8	2822	21	AAZ92646	Human alpha-1,3-fu
3	986.8	73.3	2170	21	AAZ92645	Murine alpha-1,3-f
4	939.2	69.8	2036	21	AAZ92672	Murine alpha-1,3-f
5	127.4	9.5	1814	18	AAT59505	Murine myeloid-lin
6	127	9.4	3594	18	AAT85219	Mouse alpha-fucosy
7	114.4	8.5	1701	15	AAQ77732	Human alpha-1,3-fu
8	114.4	8.5	1701	21	AAF21136	Human low adenosin
9	114.4	8.5	1701	21	AAA35014	Human adenosine re
10	114.4	8.5	6944	21	AAF21137	Human low adenosin
11	114.4	8.5	6944	21	AAA35015	Human adenosine re
12	113.4	8.4	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
13	113.4	8.4	1654	18	AAT61680	Human alpha(1,3)-f
14	112.2	8.3	1086	15	AAQ56912	Lewis blood group
15	112.2	8.3	2042	18	AAT80111	Fucosyltransferase
16	112.2	8.3	2043	12	AAQ13330	Glycosyltransferas
17	112.2	8.3	2043	15	AAQ56906	DNA encoding a gly
18	112.2	8.3	2043	18	AAT76769	Human alpha 1,3/4
19	112.2	8.3	2043	18	AAT61675	Human alpha(1,3/1,
20	111.8	8.3	1126	21	AAF21135	Human low adenosin
21	111.8	8.3	1126	21	AAA35013	Human adenosine re
22	111.8	8.3	1316	15	AAQ56911	DNA encoding a gly
23	111.8	8.3	1316	18	AAT61679	Human alpha(1,3)-f
24	103.6	7.7	795	17	AAT13798	Alpha-(1,3/1/4)-fu
25	102.4	7.6	2175	11	AAQ06691	CDX, a MILA involv
26	100.8	7.5	1256	21	AAF21134	Human low adenosin
27	100.8	7.5	1256	21	AAA35012	Human adenosine re
28	100.8	7.5	1400	13	AAQ31436	Encodes a HeLa cel
29	100.8	7.5	1488	15	AAQ56910	DNA encoding a gly
30	100.8	7.5	2175	12	AAQ14382	Clone 7.2 encoding
31	100.8	7.5	2175	24	AAS17082	Human cDNA clone 7
32	100.8	7.5	2861	12	AAQ14383	Clone 1 encoding 1
33	100.8	7.5	2861	21	AAF21133	Human low adenosin
34	100.8	7.5	2861	21	AAA35011	Human adenosine re
35	100.8	7.5	2861	24	AAS17083	Human cDNA clone 1
36	100.8	7.5	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
37	100.8	7.5	3647	15	AAQ56909	DNA encoding a gly
38	100.8	7.5	3647	18	AAT61678	Human alpha(1,3)-f
39	97.6	7.3	2134	18	AAT59506	Human myeloid deri
40	66.6	4.9	1578	23	ABL11867	Drosophila melanog
c 41	66.6	4.9	4229	23	ABL11866	Drosophila melanog
42	54.4	4.0	891	21	AAZ92670	Murine alpha-1,3-f
43	52.2	3.9	1353	21	AAZ51685	C. elegans alpha-1
44	49.8	3.7	90	21	AAC24928	Human secreted pro
c 45	44.8	3.3	2094	17	AAT33198	Alpha-1 A/D adrena

ALIGNMENTS

RESULT 1
AAZ92647
ID AAZ92647 standard; cDNA; 2676 BP.
XX
AC AAZ92647;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 295..1374
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

```
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match             100.0%;  Score 1346;  DB 21;  Length 2676;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1346;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACTGCTCTCCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29 ACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 88

Qy      61 CCCCAGTCGCGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 CCCCAGTCGCGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCG 148

Qy     121 CCGCGCAGCCCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149 CCGCGCAGCCCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGT 208

Qy     181 AATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 AATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCA 268

Qy     241 TGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCAT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 TGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCAT 328

Qy     301 TTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAAC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 TTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAAC 388

Qy     361 CTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 CTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAA 448

Qy     421 ACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 ACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGC 508

Qy     481 CATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 CATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCC 568

Qy     541 ATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     569 ATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAG 628

Qy     601 ACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 ACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGA 688

Qy     661 TTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 TTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGT 748

Qy     721 TTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     749 TTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGA 808

Qy     781 CGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     809 CGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGG 868

Qy     841 TTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     869 TTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAA 928

Qy     901 GCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 GCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGA 988

Qy     961 TTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     989 TTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATT 1048

Qy    1021 ACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1049 ACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGAC 1108

Qy    1081 CATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1109 CATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATT 1168

Qy    1141 ATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTAT 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1169 ATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTAT 1228

Qy    1201 ACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAAT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1229 ACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAAT 1288

Qy    1261 CACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTA 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1289 CACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTA 1348

Qy    1321 ATTTAGAGAAATGGTTTTGGAATTAA 1346
           |||||||||||||||||||||||||
Db    1349 ATTTAGAGAAATGGTTTTGGAATTAA 1374


RESULT 2
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
```

```
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag= a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 134-142; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

  Query Match             80.8%;  Score 1088;  DB 21;  Length 2822;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-293;
  Matches 1088;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     259 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1281 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 1340

Qy     319 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    1341 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 1400

Qy     379 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1401 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 1460

Qy     439 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1461 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 1520

Qy     499 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1521 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 1580

Qy     559 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1581 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 1640

Qy     619 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1641 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 1700

Qy     679 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1701 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 1760

Qy     739 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1761 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 1820

Qy     799 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1821 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 1880

Qy     859 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1881 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 1940

Qy     919 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1941 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 2000

Qy     979 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2001 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 2060

Qy    1039 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1098
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2061 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 2120

Qy    1099 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2121 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 2180

Qy    1159 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2181 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 2240

Qy    1219 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1278
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2241 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 2300

Qy    1279 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1338
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2301 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 2360

Qy    1339 GGAATTAA 1346
           ||||||||
Db    2361 GGAATTAA 2368


RESULT 3
AAZ92645
```

```
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag= a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match             73.3%;  Score 986.8;  DB 21;  Length 2170;
```

```
  Best Local Similarity   90.0%;  Pred. No. 2.6e-265;
  Matches 1069;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    1;

Qy    159 CCCTAGGACCGATTTAGAATGTAATAACTCAAGGATTTGATAATACAGTGAAGTAGTATA 218
          | || ||| || |||||||||| |||||||| | || |||||  || ||| ||||||  |
Db      9 CGCTGGGATCGCTTTAGAATGTGATAACTCACGCATCTGATAGCACCGTGGAGTAGTTCA 68

Qy    219 ACAACTGTCTACGTGCTTCCCATGATATGTTCTCTATATTGAAAAATTATGACATCAACA 278
            ||    |||   ||||| ||||| |||||||||| | ||||||||||||||||||||
Db     69 GCACTCATCTC--TGCTTTCCATGCTATGTTCTCTACACTGAAAAATTATGACATCAACA 126

Qy    279 TCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCA 338
          |||||||| |||||||||||||| |||| |||||||| |||||||||| |||||||||
Db    127 TCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCATCCTGGGCTGCTTCATGGCA 186

Qy    339 TGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCC 398
          ||||| |||||||| ||||| || |||||||||||| |||||||||||||||| || ||
Db    187 TGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCTTCAGTCCAATGGAGTCTGCA 246

Qy    399 AGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACT 458
          || |||||||||||||||||| ||||| ||||| |||||||||||||||| ||||||
Db    247 AGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAACTGATTATTTTAACGAAACT 306

Qy    459 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 518
          || |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    307 ACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 366

Qy    519 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
          |||||||| |||||||| |||||||||||||| |||| ||| ||||||||||||||||||
Db    367 ATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCTCATTGTACAACAAATCCCAT 426

Qy    579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 638
          || || ||||| || ||  |||||||||| |||||||||| || ||||||||||| ||
Db    427 GCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGACTAACTTACCTCAGCAGGCC 486

Qy    639 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 698
          |||||||||||| |||||||||||||||||||||||| || ||||| |||||||| ||||||
Db    487 AGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGTCACCCACTCACACCCCCCAA 546

Qy    699 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 758
          |||||||||||||| |||||||| |||||||||| ||||| ||||||||||||||||||
Db    547 AAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTTATCGCCGTGATTCAGATATC 606

Qy    759 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 818
          |||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||
Db    607 CAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCTTTGTGTTTGAAGTGCCAAGC 666

Qy    819 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 878
          || ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    667 AAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGGGTCAAG 726

Qy    879 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 938
          ||||||| ||||| ||||| || |||||||||| |||||| || |||||||| ||||||
Db    727 TATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCTATGGCCAAGCATTCGGAGAA 786

Qy    939 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 998
          || || || |||||||| ||||||| |||||||| |||||||||||||||||| |||
Db    787 TACGTGAACGATAAAAATCTGATTCCCACCATATCTACTTGTAAATTTTATCTTTCATTT 846

Qy    999 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 1058
          ||||| ||||| |||||||||||| |||||||| |||||||| ||| |||||||
Db    847 GAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCTACAATGCATTTTTGGCTGGT 906

Qy   1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
          || |||||||||| ||||| |||||||||||||||||||||||||||||||||| |||
Db    907 TCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATGAGAATTATATTCCAGCTGAT 966

Qy   1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 1178
          |||||||||||||||||||| |||||||||||||||| ||||||| ||||||||||||
Db    967 TCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGTTAGCAAAATATCTGAAGGAA 1026

Qy   1179 GTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTA 1238
          || |||||||||||||||| |||||||||||||||||||||| |||||||| ||||||
```



```
Db   1027 GTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACTGGAGAAAGGATTTTACTGTA 1086

Qy   1239 AATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCAT 1298
          || || ||||| |||||||||||||||||||  |||| ||||| ||||| |||||||||
Db   1087 AACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCATGCGACCATGTAAAAAGGCAT 1146

Qy   1299 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1346
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1147 CAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTAA 1194


RESULT 4
AAZ92672
ID   AAZ92672 standard; cDNA; 2036 BP.
XX
AC   AAZ92672;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Example 8; Page 162-164; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

  Query Match             69.8%;  Score 939.2;  DB 21;  Length 2036;
  Best Local Similarity   91.5%;  Pred. No. 5.1e-252;
  Matches  995;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy    259 GAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTA 318
          |||||||||||||||||||||||||||| |||||||||||||| |||| |||||||| |
Db      1 GAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCGTCTGCATCA 60

Qy    319 TCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCT 378
          |||||||||| |||||||||||||| |||||||| ||||| || |||||||||||| |||
Db     61 TCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACAGCTGGGTCT 120

Qy    379 TCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAAAA 438
          |||||||||||| || || || |||||||||||||||||||| ||||| ||||| ||||
Db    121 TCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCTCCACAAAAA 180

Qy    439 CTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTG 498
          |||||||||||| |||||||| |||||||| ||||| ||||||||||||||||||||||
Db    181 CTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAGACCTTTG 240

Qy    499 ACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTT 558
          |||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |
Db    241 ACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACAGACCGCT 300

Qy    559 CACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA 618
          || |||||||||||||||||| || ||||| || ||  |||||||||| ||||||||||
Db    301 CATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGGGATCTGA 360

Qy    619 CAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAAT 678
          | || |||||||||||  || |||||||||| |||||||||||||||||||||||| || |
Db    361 CTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAATTTAGAGT 420

Qy    679 CACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTT 738
          |||| |||||||| |||||||||||||||||||| |||||||| |||||||||| ||||
Db    421 CACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACTCTAACTT 480

Qy    739 ACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCT 798
          | |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    481 ATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACAAATCCCT 540

Qy    799 TCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACC 858
          | |||||||||||||||||| |||| |||||||||||||||||||||||||||||||||
Db    541 TTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAACTGGAACC 600

Qy    859 CTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCT 918
          |||||||||||| |||||||||||||| ||||| ||||| || |||||||||| ||||
Db    601 CTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCACACCT 660

Qy    919 ACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTT 978
          | || |||||||| |||||||| || || ||||||||| ||||||| ||||||||| |||
Db    661 ATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATATCTACTT 720

Qy    979 GTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT 1038
          |||||||||||||||| |||||||| ||||| ||||| |||||||||| |||||||| |
Db    721 GTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAGCTCT 780

Qy   1039 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 1098
          ||||||| ||| |||||||| || |||||||||| ||||| ||||||||||||||||||
Db    781 ACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAAAACTATG 840

Qy   1099 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 1158
          |||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||
Db    841 AGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGTGAGT 900

Qy   1159 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 1218
          ||||||| |||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db    901 TAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTACTTTAACT 960
```

```
Qy    1219 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 1278
           |||| |||||||| |||||||| || ||||| ||||||||||||||||||||  |||| |
Db     961 GGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGCCTGGCAT 1020

Qy    1279 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1338
           |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTT 1080

Qy    1339 GGAATTAA 1346
           ||||||||
Db    1081 GGAATTAA 1088


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag= a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             9.5%;  Score 127.4;  DB 18;  Length 1814;
  Best Local Similarity   52.1%;  Pred. No. 4.2e-25;
  Matches  334;  Conservative    0;  Mismatches  301;  Indels    6;  Gaps    2;

Qy     535 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 594
           | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     595 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 654
           ||||  ||  |          |   |    |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     655 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 714
             ||| | |||       |||||||  || | | | ||  |          | ||  |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     715 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 774
           | | ||  |||   |  ||||  || ||  ||||||||||||||||   || || ||  ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     775 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 834
           ||||  |     | |    ||   |   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     835 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 894
           ||||||  | || || |   |       |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     895 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 954
           |       |  |  |  ||   | ||| |  ||    |||       |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     955 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1014
           | || | || ||  |  |    |      || || ||  |||||||  || |||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    1015 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1071
           |||  |||||||| || ||| | |    ||||||  |   ||||||  |||||||  || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    1072 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1131
           ||||||||  || |||  | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    1132 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 1172
           |||| || |  | |||| || |||| || ||    |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag= a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag= b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag= c
FT                   /number= 1
FT   exon            670..1149
```

```
FT                   /*tag= d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag= e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag= f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag= g
FT                   /number= 3
FT   exon            2067..3552
FT                   /*tag= f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             9.4%;  Score 127;  DB 18;  Length 3594;
  Best Local Similarity   52.3%;  Pred. No. 7.9e-25;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     535 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 594
           | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336

Qy     595 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 654
           |||| ||  |          |   |    |||| | | |    ||||||| |   |||
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     655 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 714
           ||| | |||       |||||||| || | | | || |          | || |   |
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     715 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 774
           | | || |||  |   ||||  || || ||||||||||||||||   || || || ||
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     775 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 834
           ||||  |     | |    ||   |   |    | | ||   ||||   |   |||
Db    2517 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     835 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 894
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     895 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 954
           |       |  |  |  ||   | ||| |  ||    |||        |    | |||
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     955 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1014
           | || | || ||  |  |    |     || || ||  |||||| || |||   ||
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy    1015 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1071
           ||| ||||||||| || ||| | |    ||||||  |   ||||||  |||||| || |
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy    1072 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1131
           ||||||||  || |||  | ||| |||  || | | ||| |||| | ||  | || |
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy    1132 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
           |||| || |  | |||| || |||| || ||
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904


RESULT 7
AAQ77732
ID   AAQ77732 standard; cDNA to mRNA; 1701 BP.
XX
AC   AAQ77732;
XX
DT   22-JUN-1995  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; sialylated Lewis blood group antigen;
KW   Lewis x; Lewis a; glycosylation; glycoprotein; glycolipid; ds.
XX
OS   Homo sapiens (THP-1 cells).
XX
FH   Key             Location/Qualifiers
FT   CDS             123..1151
FT                   /*tag= a
FT                   /product= alpha-1,3-fucosyltransferase
XX
PN   WO9423021-A.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994;   94WO-JP00496.
XX
PR   29-MAR-1993;   93JP-0069016.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hanai N,  Kurata K,  Nishi T,  Sasaki K;
XX
DR   WPI; 1994-333183/41.
DR   P-PSDB; AAR63215.
XX
PT   Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
```

```
PT   of active sugar chains, e.g. sialylated Lewis X
XX
PS   Claim 3; Page 99-102; 155pp; Japanese.
XX
CC   The sequence coding for alpha-1,3-fucosyltransferase was isolated
CC   from a human THP-1 cell cDNA library. The enzyme catalyses the
CC   addition of fucose via a 1,3-linkage to N-acetylglucosamine
CC   contained in glycoproteins and glycolipids, e.g. for sialylation of
CC   Lewis blood group antigens.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             8.5%;  Score 114.4;  DB 15;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 1.8e-21;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     522 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 581
           | |  ||||    | ||||| || |   |  ||||   |||    ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     582 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 641
           ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     642 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 701
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     702 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 761
           ||   | |       | | || |||  |   ||||  ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     762 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 821
           |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     822 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 881
             |   |||    |||||| || || ||||   |      |||     | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     882 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 941
           |||    ||||           |      |  ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     942 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1001
           |    |  |       || | || ||| |  |     |      || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    1002 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1058
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
            |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973


RESULT 8
AAF21136
ID   AAF21136 standard; DNA; 1701 BP.
XX
AC   AAF21136;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2703.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1034; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1701 BP; 290 A; 527 C; 589 G; 295 T; 0 other;

  Query Match             8.5%;  Score 114.4;  DB 21;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 1.8e-21;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     522 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 581
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     582 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 641
```

```
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    642 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 701

Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    702 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 761

Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    762 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 821

Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    822 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 881

Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    882 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 941

Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    942 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1001

Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy   1002 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1058

Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy   1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118

Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy   1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162

Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973


RESULT 9
AAA35014
ID   AAA35014 standard; DNA; 1701 BP.
XX
AC   AAA35014;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2703.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 957-958; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             8.5%;  Score 114.4;  DB 21;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 1.8e-21;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy    522 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 581

Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    582 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 641

Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    642 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 701

Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    702 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 761

Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    762 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 821

Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    822 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 881

Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    882 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 941

Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    942 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1001

Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809
```

```
Qy    1002 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1058
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
           |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 10
AAF21137
ID AAF21137 standard; DNA; 6944 BP.
XX
AC AAF21137;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2704.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034-1036; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

Query Match 8.5%; Score 114.4; DB 21; Length 6944;
Best Local Similarity 50.8%; Pred. No. 3.7e-21;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy     522 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 581
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     582 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 641
           ||  |  |||| ||||| ||  |           | | |  |  | || | |       ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     642 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 701
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     702 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 761
           ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     762 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 821
           |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     822 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 881
           |   |||    |||||| || || ||||   |      |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     882 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 941
           |||    ||||           |      |  ||  ||  || |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     942 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1001
           |    |  |      || | || ||| |  |     |     || || || ||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy    1002 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1058
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy    1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
           |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy    1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

RESULT 11
AAA35015
ID AAA35015 standard; DNA; 6944 BP.
XX
AC AAA35015;
XX
DT 28-JUL-2000 (first entry)
XX

```
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2704.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 958-960; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

  Query Match             8.5%;  Score 114.4;  DB 21;  Length 6944;
  Best Local Similarity   50.8%;  Pred. No. 3.7e-21;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;

Qy     522 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 581
           | |  ||||    | |||||  || |   |  ||||    |||     ||    || | ||
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     582 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 641
           ||  |  |||| |||||  ||  |          | | |  |  | || | |       ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     642 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 701
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     702 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 761
           ||   | |       | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     762 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 821
           |||||  ||||||  | ||  |     |     ||||    |       |||||  |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     822 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 881
             |   |||    |||||| || || ||||   |      |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     882 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 941
           |||    ||||           |     |  ||  ||  || |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     942 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 1001
           |   |  |      || | || ||| |  |     |      || || || ||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy    1002 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 1058
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy    1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
           ||||  || || || ||||| ||  |  |||  | |||||||    |   | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy    1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216


RESULT 12
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
```

```
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             8.4%;  Score 113.4;  DB 15;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 3.3e-21;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
           || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
             |      ||      |  ||| |  ||||  | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
           || ||  |  | || ||||||||  |||  |   | |  |   |  |   | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
           |||      ||| || ||||           | | | |      |    | |      | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
           | |  |   |||     |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
           |      |     |     |   ||  ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043
           |||||  |||| || || ||  | |||   || ||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163
           |   | ||| |  ||  | ||||| || ||||| || |    |    ||||  || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223
             ||| ||| ||||  | |||||  || |        ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    1224 AAG 1226
           ||
Db    1095 GAG 1097
```

```
RESULT 13
AAT61680
ID   AAT61680 standard; DNA; 1654 BP.
XX
AC   AAT61680;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..1208
FT                   /*tag=  a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13643.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 6; Page 291-292; 329pp; English.
XX
CC   A DNA clone (AAT61680) codes for an enzyme (AAW13643) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC   precursors but not type I glycoconjugates.  It was identified in a
CC   human lambda phage genomic DNA library using a probe obtd. from
CC   Lewis enzyme cDNA (see also AAT61675).  The Fuc-TVI DNA can be used
CC   to construct animal cell lines with specific capabilities with
CC   respect to post-translational modification of the oligosaccharides
CC   of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC   use in oligosaccharide prodn., to genotype individuals at this
CC   fucosyltransferase locus, and in the prodn. of chimeric
CC   fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             8.4%;  Score 113.4;  DB 18;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 3.3e-21;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
           || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
             |      ||      |  ||| |  ||||  | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
           || ||  |  | || ||||||||  |||  |   | |  |   |  |   | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
```

```
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755

Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806

Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866

Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926

Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986

Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043

Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103

Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163

Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223

Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1224 AAG 1226

Db   1095 GAG 1097
```

RESULT 14
AAQ56912
ID AAQ56912 standard; DNA; 1086 BP.
XX
AC AAQ56912;
XX
DT 26-JUL-1994 (first entry)
XX
DE Lewis blood group fucosyltransferase DNA.
XX
KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW surface; oligosaccharide; ss.
XX
PN WO9402616-A.
XX
PD 03-FEB-1994.
XX
PF 20-JUL-1993; 93WO-US06703.
XX
PR 20-JUL-1992; 92US-0914281.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Lowe JB;
XX
DR WPI; 1994-048874/06.
XX
PT DNA fragment encoding a glycosyltransferase - can be used for in
PT vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT blood gp. determinants, to protect against transplant rejection
XX
PS Disclosure; Fig 6; 249pp; English.
XX
CC The sequence is that encoding a Lewis blood group fucosyl
CC transferase (Fuc-TIII). The enzyme produced by the DNA may be
CC non glycosylated. This prevents premature loss of enzyme activity.
CC It can also be used in in vitro reactions to modify cell surface
CC oligosaccharide mols. e.g. blood group determinants.
CC See also AAQ56905-11.
XX
SQ Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

Query Match 8.3%; Score 112.2; DB 15; Length 1086;
Best Local Similarity 51.6%; Pred. No. 5.7e-21;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596

Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653

Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713

Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773

Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884

Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944

Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004

Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061

Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121

Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181

Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226

Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972
```

RESULT 15
AAT80111
ID AAT80111 standard; cDNA; 2042 BP.
XX
AC AAT80111;
XX
DT 01-DEC-1997 (first entry)

XX
DE Fucosyltransferase III coding sequence.
XX
KW P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW adult respiratory distress syndrome; tumour metastasis;
KW fucosyltransferase III; acetylglucosaminyltransferase;
KW autoimmune disorder; inflammatory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO9706176-A2.
XX
PD 20-FEB-1997.
XX
PF 02-AUG-1996; 96WO-US12820.
XX
PR 17-MAY-1996; 96US-0649802.
PR 03-AUG-1995; 95US-0510920.
PR 15-MAY-1996; 96US-0017794.
XX
PA (OKLA) UNIV OKLAHOMA STATE.
XX
PI Cummings RD, Mcever RP, Moore KL;
XX
DR WPI; 1997-154206/14.
XX
PT New O-glycan(s) derived from P-selectin glyco:protein ligand
PT (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT for treating inflammatory or autoimmune disorders or tumours
XX
PS Disclosure; Page 80-81; 96pp; English.
XX
CC This sequence represents the coding sequence for the human
CC fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC protein expressed on the surfaces of activated platelets and endothelium
CC in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC high affinity P-selectin ligand produced by leukocytes. Binding of
CC P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC highly extended extracellular domain, which is a feature of mucin-like
CC proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC O-glycans, and O-glycanated products of the invention. The products are
CC used for inhibiting the binding of P-selectin and other selectins to
CC PSGL-1. They can be used for the treatment of inflammation, injury
CC resulting from ischaemia and reperfusion, bacterial sepsis and
CC disseminated intravascular coagulation, adult respiratory distress
CC syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC other autoimmune or inflammatory disorders. The products can also be used
CC in the production of antibodies and in diagnostic applications.
XX
SQ Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

Query Match 8.3%; Score 112.2; DB 18; Length 2042;
Best Local Similarity 51.6%; Pred. No. 8e-21;
Matches 364; Conservative 0; Mismatches 323; Indels 18; Gaps 4;

```
Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
            | ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
```

```
          |||||  |||  |  ||  |||||  ||||  ||  ||    |  ||    |   | | |  ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
            |  ||  ||  ||  ||  ||  ||  |  ||||||  |  ||  ||  ||  |||      |||  ||  |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |     |  |     |     |           |  |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
            |  |||||  |||||  |||   ||||||||  ||  ||    |||||  ||    ||  |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||          |  |  |  |  |||||  |    |                 |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          ||  |  |  |  ||||    ||  | ||     |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
           ||  | |||   || ||||||||  || |||||  |     ||  ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
           ||  ||  ||  ||  |||||  ||    ||    ||||| ||||  |   |  ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
           |||||  ||  |||||  ||  |   |    ||||  ||  ||  ||    |||  |||  ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
           |||||  ||  |       |||||  ||  ||||||  ||||  ||  ||
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 5, 2002, 23:16:51
Job time : 182.049 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 1126.97 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-5_COPY_29_1374
Perfect score: 1346
Sequence: 1 actgctctcccctgcagctc.....agaaatggttttggaattaa 1346

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	573.8	42.6	697	9	BB397507	BB397507 BB397507
2	504.2	37.5	568	10	BM021317	BM021317 ie75f03.y
3	492.8	36.6	607	10	BI102760	BI102760 602888338
4	422	31.4	716	9	AV725071	AV725071 AV725071
5	387.2	28.8	648	9	BB595164	BB595164 BB595164
6	375	27.9	625	9	BB649967	BB649967 BB649967
7	363.2	27.0	767	9	BB640911	BB640911 BB640911
8	324.8	24.1	649	9	AV338075	AV338075 AV338075
9	322.2	23.9	628	9	BB651409	BB651409 BB651409
c 10	307.4	22.8	579	10	BM021081	BM021081 ie75f03.x
c 11	296.8	22.1	1101	12	CNS059ST	AL327638 Tetraodon
12	239.2	17.8	670	9	BB626126	BB626126 BB626126
13	212.2	15.8	533	10	BI604066	BI604066 603243886
c 14	184.2	13.7	570	10	BM082484	BM082484 fu25e08.x
c 15	177	13.2	1100	12	CNS0579Z	AL324368 Tetraodon
c 16	170.2	12.6	1016	12	CNS0429P	AL271222 Tetraodon
c 17	169.4	12.6	335	12	AQ054539	AQ054539 CIT-HSP-2
18	169	12.6	617	10	BE266935	BE266935 601189956
c 19	165.2	12.3	487	12	AQ246164	AQ246164 HS_2060_A
20	165	12.3	844	12	CNS01V8V	AL168808 Tetraodon
c 21	157.4	11.7	887	12	CNS02Q0S	AL208693 Tetraodon
22	155.4	11.5	651	10	BJ036855	BJ036855 BJ036855
c 23	151.6	11.3	666	10	BJ050300	BJ050300 BJ050300
c 24	151.2	11.2	711	12	CNS034XF	AL228012 Tetraodon
c 25	151	11.2	501	12	B41073	B41073 HS-1052-B2-
26	149.6	11.1	367	9	BB843869	BB843869 BB843869
27	146.6	10.9	640	10	BJ073227	BJ073227 BJ073227
28	136.2	10.1	199	9	BB398103	BB398103 BB398103
c 29	123.4	9.2	674	10	BE605503	BE605503 fl10h04.x
30	120.6	9.0	971	12	CNS047IV	AL278032 Tetraodon
c 31	120.4	8.9	615	10	BI326811	BI326811 STR00078
c 32	117.8	8.8	1023	12	CNS02ANG	AL188773 Tetraodon
33	114.8	8.5	661	9	BB651647	BB651647 BB651647
34	102.8	7.6	487	10	BE863396	BE863396 UI-M-BH0-
c 35	100.2	7.4	702	10	BJ043578	BJ043578 BJ043578
36	93.8	7.0	863	12	CNS02ROF	AL210840 Tetraodon
37	90.4	6.7	720	10	BF180380	BF180380 601804564
38	90	6.7	551	10	BJ035701	BJ035701 BJ035701
39	87.4	6.5	601	10	BJ036679	BJ036679 BJ036679
40	87.4	6.5	605	10	BJ028804	BJ028804 BJ028804
c 41	87	6.5	1056	12	CNS01VL7	AL169252 Tetraodon
42	86.8	6.4	304	9	AU081478	AU081478 AU081478
c 43	84.6	6.3	102	12	B70664	B70664 CIT-HSP-206
44	84.2	6.3	638	10	BJ035697	BJ035697 BJ035697
45	83.6	6.2	1057	12	CNS03ADC	AL235065 Tetraodon

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN
```

Query Match 42.6%; Score 573.8; DB 9; Length 697;
Best Local Similarity 91.4%; Pred. No. 2.2e-133;
Matches 619; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

```
Qy    670 ATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGA 729
          | || || ||||| |||||||| ||| |||||||||||| ||| |||||||| |||||||
Db      4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy    730 CTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCA 789
          |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy    790 CAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTA 849
          |||||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||
Db    123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy    850 ACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAA 909
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||
Db    183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy    910 TCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCA 969
          |||| ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||
Db    243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy    970 TATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGG 1029
          ||||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||||
Db    303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy   1030 AAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGG 1089
          ||||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db    363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy   1090 AAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTC 1149
          |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||
Db    423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy   1150 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTT 1209
          |||||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db    483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy   1210 ACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCAT 1269
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||
Db    543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy   1270 GTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 1329
          |  |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db    603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy   1330 AATGGTTTTGGAATTAA 1346
          ||||||||||||||||
Db    663 AATGGTTTTGGAATTAA 679
```

```
RESULT 2
BM021317
LOCUS       BM021317                568 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021317
VERSION     BM021317.1  GI:16535673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ie75f03.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
```

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 127 a 170 c 123 g 148 t
ORIGIN

Query Match 37.5%; Score 504.2; DB 10; Length 568;
Best Local Similarity 99.4%; Pred. No. 6.5e-116;
Matches 506; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 ACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 ACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 119

Qy      61 CCCCAGTCGCGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 CCCCAGTCGCGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCG 179

Qy     121 CCGCGCAGCCCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGT 180
           |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db     180 CCGCGCAGCCCTCGCGAGCGTCTTGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGT 239

Qy     181 AATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 AATAACTCAAGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCA 299

Qy     241 TGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCAT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 TGATATGTTCTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCAT 359

Qy     301 TTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAAC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 TTTTAATTGTCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAAC 419

Qy     361 CTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 CTACCAACAGCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAA 479

Qy     421 ACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACTTCTTTTCCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGC 539

Qy     481 CATTTGGGCAGACCTTTGACCTTACATCC 509
           |||||||||||||||||||||||||||||
Db     540 CATTTGGGCAGACCTTTGACCTTACATCC 568
```

RESULT 3
BI102760
LOCUS BI102760 607 bp mRNA linear EST 26-JUN-2001
DEFINITION 602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609 5', mRNA sequence.
ACCESSION BI102760
VERSION BI102760.1 GI:14553653
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 607)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11119 row: l column: 10
High quality sequence stop: 606.

FEATURES Location/Qualifiers
source 1. .607
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5043609"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT 179 a 149 c 121 g 157 t 1 others
ORIGIN

Query Match 36.6%; Score 492.8; DB 10; Length 607;
Best Local Similarity 90.1%; Pred. No. 4.9e-113;
Matches 549; Conservative 0; Mismatches 58; Indels 2; Gaps 2;

```
Qy     489 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACA 548
           |||||||||| |||||||||||||||||||||||||| |||||||| ||||||||||||
Db       1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy     549 ACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGT 608
           || ||||| ||| |||||||||||||||||| || ||||| || ||  ||||||||||
Db      61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy     609 TGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATG 668
           |||||||||| || |||||||||| || |||||||||| |||||||||| |||||||
Db     120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy     669 AATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTG 728
           ||||| || ||||| |||||||| |||||||||||||||||||| |||||||| ||||||
Db     179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy     729 ACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGC 788
           ||||| ||||| |||||||||||||||||||||||||||||||||||||||||||| |||
Db     239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy     789 ACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGT 848
           |||||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||
Db     299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy     849 AACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAA 908
           |||||||||||||||||||| |||||||||||||| ||||| ||||| || ||||||
Db     359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy     909 ATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACC 968
           ||||| ||||| || |||||||| |||||||| || || |||||||| ||||||| |||
Db     419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy     969 ATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACG 1028
           |||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||
Db     479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy    1029 GAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGG 1088
           |||||||| |||||||| ||| ||||||| || |||||| ||| ||||| |||||||||
Db     539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy    1089 GAAAACTAT 1097
           |||||||||
```

```
Db    599 GAAAACTAT 607


RESULT 4
AV725071
LOCUS       AV725071                716 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION   AV725071
VERSION     AV725071.1  GI:10830091
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .716
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCBLB02"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      219 a    113 c    133 g    249 t      2 others
ORIGIN

  Query Match             31.4%;  Score 422;  DB 9;  Length 716;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-95;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     925 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 984
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy     985 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 1044
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy    1045 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 1104
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy    1105 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 1164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy    1165 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 1224
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy    1225 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 1284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy    1285 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 1344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy    1345 AA 1346
           ||
Db     421 AA 422


RESULT 5
BB595164
LOCUS       BB595164                648 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB595164 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030004N01 5', mRNA sequence.
ACCESSION   BB595164
VERSION     BB595164.2  GI:16450167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11491766.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
```

/clone="C030004N01"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 141 a 206 c 138 g 163 t
ORIGIN

Query Match 28.8%; Score 387.2; DB 9; Length 648;
Best Local Similarity 80.4%; Pred. No. 1.7e-86;
Matches 479; Conservative 0; Mismatches 113; Indels 4; Gaps 2;

```
Qy      10 CCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCG 69
           ||| |   ||     | || |||||||| ||||  | | | |  ||||   | || |
Db      57 CCCGGTGCCTTGAGACCCCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTA 116

Qy      70 CGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGC 129
           ||||  ||||||||||||||||||| ||||| | ||  |  | || || ||||||||| ||
Db     117 CGCTACTAGGACAGCGCCGCCACCACCGCCAGCCCTCG--TCCCGCCCGCGCCGCGCTGC 174

Qy     130 CCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCA 189
           |||| || || |        ||  |    | || ||| || ||||||||||| ||||||||
Db     175 CCTCCCGCGCCCTGGATGGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCA 234

Qy     190 AGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTT 249
           | || |||||  || ||| ||||||  | ||    |||   ||||| ||||| |||||||
Db     235 CGCATCTGATAGCACCGTGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTT 292

Qy     250 CTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTG 309
           ||||| | |||||||||||||||||||||||||||||| |||||||||||||| |||| |
Db     293 CTCTACACTGAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCG 352

Qy     310 TCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACA 369
           ||||||| |||||||||| |||||||||||||| |||||||| ||||| || |||||||
Db     353 TCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACA 412

Qy     370 GCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTT 429
           ||||| |||||||||||||| || || || |||||||||||||||||| ||||| |
Db     413 GCTGGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCT 472

Qy     430 CCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGC 489
           |||| |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||
Db     473 CCACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGC 532

Qy     490 AGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAA 549
           |||||||||||||||||||||||||||||||||||||||| |||| ||| ||||||||||||
Db     533 AGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCCAGGGTGCCATCTCACAA 592

Qy     550 CGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATC 605
           | ||||| ||| |||||||||||||||||| || ||||| || ||   |||||||
Db     593 CAGACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATATAGACATC 648
```

RESULT 6
BB649967
LOCUS BB649967 625 bp mRNA linear EST 26-OCT-2001
DEFINITION BB649967 RIKEN full-length enriched, 16 days embryo head Mus musculus cDNA clone C130099N17 5', mRNA sequence.
ACCESSION BB649967
VERSION BB649967.1 GI:16484222
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 625)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .625
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C130099N17"
/clone_lib="RIKEN full-length enriched, 16 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT 136 a 198 c 135 g 156 t
ORIGIN

Query Match 27.9%; Score 375; DB 9; Length 625;
Best Local Similarity 80.9%; Pred. No. 2e-83;
Matches 462; Conservative 0; Mismatches 105; Indels 4; Gaps 2;

```
Qy      10 CCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCG 69
           ||| |   ||     | || |||||||| ||||  | | | |  ||||   | || |
Db      58 CCCGGTGCCTTGAGACCCCGCCGCCGCTCTCGCCTCTTGGCTACCCCCCGGACTCCGGCTA 117

Qy      70 CGCTCTTAGGACAGCGCCGCCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCCGCAGC 129
           |||| ||||||||||||||||||| ||||| | ||  |  | || || ||||||||| ||
Db     118 CGCTACTAGGACAGCGCCGCCACCACCGCCAGCCCTCG--TCCCGCCCGCGCCGCGCTGC 175

Qy     130 CCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCA 189
           |||| || || |         ||  |    | || ||| || ||||||||||| ||||||||
Db     176 CCTCCCGCGCCCTGGATGGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCA 235

Qy     190 AGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTT 249
           | || |||||  || ||| ||||||  | ||    |||   ||||| ||||| ||||||
Db     236 CGCATCTGATAGCACCGTGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTT 293

Qy     250 CTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTG 309
           ||||| | ||||||||||||||||||||||||||||| ||||||||||||||| |||| |
Db     294 CTCTACACTGAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCG 353

Qy     310 TCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACA 369
           ||||||| |||||||||| |||||||||||||| |||||||| ||||| || |||||||
Db     354 TCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACA 413

Qy     370 GCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTT 429
           ||||| |||||||||||||||| || || || |||||||||||||||||||| ||||| |
Db     414 GCTGGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCT 473

Qy     430 CCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGC 489
           |||| |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||
Db     474 CCACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGC 533

Qy     490 AGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAA 549
           |||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db     534 AGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAA 593

Qy     550 CGGACCGTTCACTGTACAACAAATCCCATGC 580
           | ||||| ||| ||||||||||||||||||
Db     594 CAGACCGCTCATTGTACAACAAATCCCATGC 624
```

RESULT 7
BB640911
LOCUS BB640911 767 bp mRNA linear EST 26-OCT-2001
DEFINITION BB640911 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730093G08 5', mRNA sequence.
ACCESSION BB640911
VERSION BB640911.1 GI:16476069
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 767)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .767
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730093G08"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 175 a 239 c 176 g 172 t 5 others
ORIGIN

Query Match 27.0%; Score 363.2; DB 9; Length 767;

```
  Best Local Similarity   78.8%;  Pred. No. 2e-80;
  Matches  458;  Conservative    0;  Mismatches  119;  Indels    4;  Gaps    2;

Qy      10 CCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCG 69
           ||| |   ||     | || |||||||| ||||  | | | |  ||||   | || |
Db      94 CCCGGTGCCTTGAGACCCCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTA 153

Qy      70 CGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGC 129
           ||||  |||||||||||||||||| ||||| | ||  |  | || || |||||||| ||
Db     154 CGCTACTAGGACAGCGCCGCCACCACCGCCAGCCCTCG--TCCCGCCCGCGCCGCGCTGC 211

Qy     130 CCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCA 189
           |||| || || |        ||  |    | || ||| || |||| ||||| ||||||||
Db     212 CCTCCCGCGCCCTGGATGGAGCTTTCCCTCGCTGGGATCGCTTTANAATGTGATAACTCA 271

Qy     190 AGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTT 249
           | || |||||  || ||| ||||||  | ||    |||   ||||| ||||| ||||||
Db     272 CGCATCTGATAGCACCGTGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTT 329

Qy     250 CTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTG 309
           ||||| | ||||||||||||||||||||||||||||| ||||||||||||||| |||| |
Db     330 CTCTACACTGAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCG 389

Qy     310 TCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACA 369
           ||||||| |||||||||| |||||||||||||| |||||||| ||||| || |||||||
Db     390 TCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACA 449

Qy     370 GCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTT 429
           ||||| |||||||||||||||| || || || |||| |||||||||||||| ||||| |
Db     450 GCTGGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGGGCTGAAAATGAAAAATTTCTTCT 509

Qy     430 CCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGC 489
           |||| |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||
Db     510 CCACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGC 569

Qy     490 AGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAA 549
           |||||||||||||||||| |||||||||||||| ||||| ||  |||||| ||||
Db     570 AGACCTTTGACCTTACATCCGGCCAAGCAATGTTCAATATCCAGGGGGGCCATCTTACAA 629

Qy     550 CGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATC 590
           |  | || ||| || ||||||||||||| |  |  ||||||
Db     630 CAAAACGGTCATTGGACAACAAATCCCAGGGGGGCCTGATC 670


RESULT 8
AV338075
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6378127.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .649
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="6430407H24"
                      /clone_lib="RIKEN full-length enriched, adult male
                      olfactory bulb"
                      /sex="male"
                      /tissue_type="olfactory brain"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 100.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                      3']. cDNA was cloned into the XhoI and BamHI sites.
                      Vector: a modified pBluescript KS(+) after bulk excision
                      from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                      BamHI"
BASE COUNT      209 a    105 c    117 g    218 t
ORIGIN

  Query Match             24.1%;  Score 324.8;  DB 9;  Length 649;
  Best Local Similarity   91.5%;  Pred. No. 8.5e-71;
  Matches  344;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy     971 ATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGA 1030
           |||  ||||||| |||  |||||| |||||||| ||||| ||||||||||||||||| ||
Db       2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61
```

```
Qy   1031 AAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGA 1090
          |||||| ||||||||| ||| ||||||| || |||||||||||| ||||| ||||||||||
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy   1091 AAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCC 1150
          |||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy   1151 CAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTA 1210
          ||||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy   1211 CTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATG 1270
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy   1271 TTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 1330
            |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy   1331 ATGGTTTTGGAATTAA 1346
          ||||||||||||||||
Db    362 ATGGTTTTGGAATTAA 377
```

RESULT 9
BB651409

```
LOCUS       BB651409                628 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB651409 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230079N02 5', mRNA sequence.
ACCESSION   BB651409
VERSION     BB651409.1  GI:16485662
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C230079N02"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 479.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      129 a    200 c    143 g    152 t      4 others
ORIGIN
```

```
  Query Match           23.9%;  Score 322.2;  DB 9;  Length 628;
  Best Local Similarity   78.1%;  Pred. No. 3.8e-70;
  Matches  411;  Conservative    0;  Mismatches  111;  Indels    4;  Gaps    2;

Qy      10 CCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCG 69
           ||| |   ||     | || |||||||| |||| | | | |  ||||   | || |
Db      94 CCCGGTGCCTTGAGACCCCGCCGCCGCTCTCGCCTCTTGGCTACCCCCCGGACTCCGGCTA 153

Qy      70 CGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGC 129
           ||||  |||||||||||||||||| ||||| | ||  |  | || || ||||||||| ||
Db     154 CGCTACTAGGACAGCGCCGCCACCACCGCCAGCCCTCG--TCCCGCCCGCGCCGCGCTGC 211

Qy     130 CCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCA 189
           |||| || || |         ||  |    | || ||| || |||||||||| ||||||||
Db     212 CCTCCCGCGCCCTGGATGGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCA 271

Qy     190 AGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTT 249
           | || |||||  || ||| ||||||  | ||    |||   ||||| ||||| ||||||
Db     272 CGCATCTGATAGCACCGTGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTT 329

Qy     250 CTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTG 309
           ||||| | ||||||||||||||||||||||||||||||| ||||||||||||||| |||| |
Db     330 CTCTACACTGAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCG 389

Qy     310 TCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACA 369
           ||||||| |||||||||| |||||||||||||| |||||||| ||||| || |||||||
Db     390 TCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACA 449

Qy     370 GCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTT 429
           ||||| |||||||||||||| || || || |||||||||||||||||||| ||||| |
Db     450 GCTGGGTCTTCAGTCCAATGGAGTCTGCAAGTTCTGTGCTGAAAATGAAAAATTTCTTCT 509

Qy     430 CCACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGC 489
           |||| |||||||||||| |||| |||||||| |||||||  | ||| ||||||| |||||
```

```
Db    510 CCACAAAAACTGATTATGTTAACGAAACTACCATTCTGGNTTAGGTATGGCCATNTGGGC 569

Qy    490 AGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGG 535
          |||||| |||||||||||||||||   |||||  | ||||||||
Db    570 AGACCTNTGACCTTACATCCTGCCAAAGCATGTTTCATATCCAAGG 615


RESULT 10
BM021081/c
LOCUS       BM021081                579 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;,
            mRNA sequence.
ACCESSION   BM021081
VERSION     BM021081.1  GI:16535437
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT      203 a    103 c     93 g    180 t
ORIGIN

  Query Match            22.8%;  Score 307.4;  DB 10;  Length 579;
  Best Local Similarity  99.7%;  Pred. No. 1.9e-66;
  Matches  308;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1038 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 1097
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy   1098 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 1157
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy   1158 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 1217
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy   1218 TGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCT 1277
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy   1278 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 1337
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy   1338 TGGAATTAA 1346
          |||||||||
Db    279 TGGAATTAA 271


RESULT 11
CNS059ST/c
LOCUS       CNS059ST                1101 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            045P22 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL327638
VERSION     AL327638.1  GI:8221227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="045P22"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA045DH11A1~end : T3"
BASE COUNT      287 a    249 c    269 g    278 t     18 others
ORIGIN

  Query Match            22.1%;  Score 296.8;  DB 12;  Length 1101;
  Best Local Similarity  58.3%;  Pred. No. 1.1e-63;
```

```
  Matches  603;  Conservative    0;  Mismatches  423;  Indels    9;  Gaps    5;

Qy     316 TTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGA 375
           | ||| |||| || |   || |         | || |||   || ||  ||| |||||
Db    1064 TCATCGTGGGGTGCTNTGTGACTGTGTNNTTTATGTACTNTAATCCATCCACCAGCT-GN 1006

Qy     376 TCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCA 435
           | | | ||||  |||| |||          |  | |  | || ||     | || ||||
Db    1005 TTTCCGGTCCTGTGGAGGCAGGACTGAGCTCGGAGCATGTCAAGAATCAACTATCAACCA 946

Qy     436 AAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCT 495
           | | |||| |    ||    || | ||| ||| ||||| | ||||| || || || ||||
Db     945 AGAGTGATAAA---AACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCT 889

Qy     496 TTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACC 555
           | ||||| |     ||| | ||    |||  | |  | || ||  |  | ||  |||||
Db     888 TCGACCTGAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACA 829

Qy     556 GTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATC 615
           |    || |  ||||| ||  |||  ||| |  |||| ||| ||||||||    | || |
Db     828 GAAACCTTTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACC 769

Qy     616 TGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGG 675
           || | ||  | ||        |   | ||  || ||||||||||||| ||| |  |  |||
Db     768 TGTCCAACATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGG 709

Qy     676 AATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGA 735
           | || ||  | ||| |  |      |   ||| || |||  | |||| ||||||||| ||||
Db     708 AGTCTCCCTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGA 649

Qy     736 CTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATC 795
             || || |  ||||| || ||        ||  |   || |      ||   |||  ||
Db     648 ACTATCGTCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATG 589

Qy     796 CCTTCGTG---TTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 852
           |    | |  ||||      || |||||| | ||| |||| || ||| |||| || ||||
Db     588 CTGAGGAGGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACT 529

Qy     853 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 912
           ||||||  || ||||| |||| || |  |||| ||  |  |||      ||||  |||
Db     528 GGAACCAGGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCC 469

Qy     913 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 972
           | || || || | |||||| ||  || |  ||   ||| || |      | || |||||
Db     468 ACACGTATGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCG 409

Qy     973 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 1032
           |    || || || || || || ||||| || || |||||||| || ||| | || ||||
Db     408 CCAGCTGCAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAA 349

Qy    1033 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 1092
           | || |||||  |  ||   |  ||  | || || || |||||||| ||| || |||| |.
Db     348 AACTGTACAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGA 289

Qy    1093 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 1152
           |||| |||||    || |  | |||| | || || || |||||||| | |  |||  |||
Db     288 ACTACGAGAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCA 229

Qy    1153 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 1212
             ||||| ||  | ||| |  |   | ||||||||  |  |  | |||||||  || |
Db     228 AGGAGCTTGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATT 169

Qy    1213 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 1272
           || |   |  ||||    |        ||        |  ||||       |   | |||
Db     168 TTCANGTGGCGAAGACACTTTAAAGCAAGGGCAGTGGCATTTGCCCGAGGC-CACCTGTC 110

Qy    1273 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAA 1331
           |||||| ||| | || |  |||| | |  |||| | ||| |   |  | ||| | ||
Db     109 AGGCTTGTGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAG 50

Qy    1332 TGGTTTTGGAATTAA 1346
           ||||  |||  || |
Db      49 TGGTACTGGGGTTGA 35
```

```
RESULT 12
BB626126
LOCUS       BB626126                670 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB626126 RIKEN full-length enriched, adult male diencephalon Mus
            musculus cDNA clone 9330161D06 5', mRNA sequence.
ACCESSION   BB626126
VERSION     BB626126.1  GI:16464204
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .670
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9330161D06"
                     /clone_lib="RIKEN full-length enriched, adult male
                     diencephalon"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
```

Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 107 a 246 c 147 g 169 t 1 others
ORIGIN

Query Match 17.8%; Score 239.2; DB 9; Length 670;
Best Local Similarity 76.5%; Pred. No. 2.8e-49;
Matches 319; Conservative 0; Mismatches 94; Indels 4; Gaps 2;

```
Qy     10 CCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAGCCCCAGTCG 69
Db    258 CCCGGTGCCTTGAGACCCCGCCGCCGCTCTCGCCTCTTGGCTACCCCCGGACTCCGGCTA 317

Qy     70 CGCTCTTAGGACAGCGCCGCCACCGCCGCCTGGCCCTGCCTGCCTCCTGCGCCGCGCAGC 129
Db    318 CGCTACTAGGACAGCGCCGCCACCACCGCCAGCCCTCG--TCCCGCCCGCGCCGCGCTGC 375

Qy    130 CCTCGCGAGCGCCCCGGATGGCGCTTTACCCCTAGGACCGATTTAGAATGTAATAACTCA 189
Db    376 CCTCCCGCGCCCTGGATGGAGCTTTCCCTCGCTGGGATCGCTTTAGAATGTGATAACTCA 435

Qy    190 AGGATTTGATAATACAGTGAAGTAGTATAACAACTGTCTACGTGCTTCCCATGATATGTT 249
Db    436 CGCATCTGATAGCACCGTGGAGTAGTTCAGCACTCATCT--CTGCTTTCCATGCTATGTT 493

Qy    250 CTCTATATTGAAAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTG 309
Db    494 CTCTACACTGAAAAATTATGACATCAACATCCAAAGGCATTCTTCGCCCATTTCTAATCG 553

Qy    310 TCTGCATTATCCTGGGCTGTTTCATGGCATGTCTTCTCATTTACATCAAACCTACCAACA 369
Db    554 TCTGCATCATCCTGGGCTGCTTCATGGCATGTCTGCTCATTTATATCAAGCCCACCAACA 613

Qy    370 GCTGGATCTTCAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCT 426
Db    614 GCTGGGTCTTCAGTCCAATGGAGTCTGNCAGTTCTGTGCTGAAAATGAAAAATTTCT 670
```

RESULT 13
BI604066
LOCUS BI604066 533 bp mRNA linear EST 07-SEP-2001
DEFINITION 603243886F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5286309 5', mRNA sequence.
ACCESSION BI604066
VERSION BI604066.1 GI:15497006
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 533)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11723 row: l column: 22
High quality sequence stop: 526.
FEATURES Location/Qualifiers
source 1. .533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5286309"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 115 a 172 c 119 g 127 t
ORIGIN

Query Match 15.8%; Score 212.2; DB 10; Length 533;
Best Local Similarity 78.0%; Pred. No. 1.6e-42;
Matches 372; Conservative 0; Mismatches 63; Indels 42; Gaps 8;

```
Qy      1 ACTGCTCTCCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 60
Db     58 ACTGCTCTCCCCTGCAGCTCCCCGCGCCCCCGCCGCTGTCGCTGCCTCGGTGTCCCCCAG 117

Qy     61 CCCCAGTCGCGCTCTTAGGACA-GCGCCGCCACCGCCGCCTGGCCCTGCC---------- 109
Db    118 CCCCAGTCGCGCTCTTAGGACATGCGCCGCCACCGCCGCCTGGCCCTGCCTTGCCTTCCT 177

Qy    110 --------------TGCCTCCTGCGCCGCGCAGCCCTCGCGAGCGCCCCGGATGGCGCTT 155
Db    178 TGCTGCCCGCTGCATGCCCCTTCTGGCGAAGCTGCCCCGGGAAAGGGCGGCTTTTAACCC 237

Qy    156 TACCCCTAGGACCGATTT-------AGAATGTAATAACTCAAGGATTTGATAATA----C 204
Db    238 CCCTTAAGGACCCGACTTTACGAAATGTAATTAAACTCAAACGGCATTTCGAATAATACA 297

Qy    205 AGTGAAGTAGTATAACAACTGTC-TACGTGCTTCCCATGATATGT---TCTCTATATTGA 260
Db    298 CGTGAAGTAGTATAACAACTGTCGTACGTGCTTCCCATGATATGTTCGTCTATATTGCAA 357

Qy    261 AAAATTATGACATCAACATCCAAAGGAATTCTTCGCCCATTTTTAATTGTCTGCATTATC 320
Db    358 AAAGTTATGACATCAACATCCAAAGGAATTCTTCGCCCA-TTTGGATTGTCTGCATTATC 416

Qy    321 CTGGGCTGTTTCATGG-CATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTT 379
Db    417 CTGGGCTGTTTCATGGCCATGTCTTCTCATTTACATCAAACCTACCAACAGCTGGATCTT 476

Qy    380 CAGTCCAATGGAATCAGCCAGCTCTGTGCTGAAAATGAAAAACTTCTTTTCCACCAA 436
Db    477 CAGTCCAATGGAATCAGCCAGCTCTGTGCTGAACATGACAAACTTCTTTTCCACCAA 533
```

RESULT 14
BM082484/c
LOCUS BM082484 570 bp mRNA linear EST 14-NOV-2001
DEFINITION fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223 3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;, mRNA sequence.
ACCESSION BM082484
VERSION BM082484.1 GI:16929414
KEYWORDS EST.
SOURCE zebrafish.

```
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN

  Query Match             13.7%;  Score 184.2;  DB 10;  Length 570;
  Best Local Similarity   59.2%;  Pred. No. 1.8e-35;
  Matches  337;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    1;

Qy     648 TTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGC 707
           |||||||| |||||||||  |  |||| |||||  || ||  |||||     ||   |||
Db     570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy     708 ATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCT 767
           | ||  || |||| ||||  ||||| |  ||| | |  ||| | ||||| | ||| |
Db     510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy     768 TATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAA 827
           | ||    ||    |  | |    ||   |  ||||   | || |  || || |||
Db     450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy     828 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGAGTCAAG 878
           | || || ||| ||||||||||||      |||          ||      || ||   || |
Db     390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331

Qy     879 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 938
           |||    ||  |||||  |  |||  |  | ||    | |||| | || | |  |  |
Db     330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271

Qy     939 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 998
           ||| |  | ||||| | |    |||||| |  ||   || |||||||| || |||||
Db     270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211

Qy     999 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 1058
           || || || |||||||||| |||||||| || ||| |  |    | || ||   || ||
Db     210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151

Qy    1059 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 1118
           |||| ||||| ||| |||| ||  | ||| ||||||||| || |  ||||| | |||
Db     150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91

Qy    1119 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 1178
           | || |||||||| | || |       | |||    ||||||   ||| || |
Db      90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31

Qy    1179 GTCGACAAAAACAATAAGTTATACCTTAG 1207
           |  | |||||  | |    ||| |  ||
Db      30 CTAAATAAAAATGACATTGCATATCACAG 2


RESULT 15
CNS0579Z/c
LOCUS       CNS0579Z                1100 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            007H05 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL324368
VERSION     AL324368.1  GI:9557250
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007H05"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA007CD03C1~end : T7"
BASE COUNT      314 a    228 c    252 g    304 t      2 others
ORIGIN

  Query Match             13.2%;  Score 177;  DB 12;  Length 1100;
  Best Local Similarity   55.4%;  Pred. No. 1.5e-33;
  Matches  428;  Conservative    0;  Mismatches  332;  Indels   13;  Gaps    4;

Qy     462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
           || ||| | ||| | ||||| | ||| ||||  ||||         || ||  |||
Db     773 ATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTTGCATACGACTCTTGTAAAGTTTAC 714
```

```
Qy    522 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 581
          | ||| ||  | || ||  | ||||||   || || || || ||| ||||| | || ||
Db    713 TACAATATTGACGGCTGTGAGCTCACAGATGATCGATCCCTTTACGACAAAGCGCAGGCT 654

Qy    582 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 641
          |||||| ||   |||   |||||    ||| |||||  ||| || ||   | |  |  ||
Db    653 GTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGGAAACTTGCCAGTGGAGCCGCGG 594

Qy    642 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 701
          |||   || ||||  |||||||||| |  |  |||||||| || |   ||||   |   |
Db    593 CCATATTTTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGATA 534

Qy    702 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 761
             ||| | ||     | || ||| |||||||||  ||| |    ||||| || ||   |
Db    533 CCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGTA 474

Qy    762 GTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 818
          |     ||   |     | | |  |     | |    |         | |   ||||
Db    473 GCCAGGTACCCTCTTACCATCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGAG 414

Qy    819 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC------ATGCCAGA 872
          ||| | ||| | |  || ||| | || || ||    |   |            | | |||
Db    413 AAAAACAAACTTGCCTGTTGGATCGTTAGCAATGCTACATCCACTGGGACAGGTACAAGA 354

Qy    873 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 932
          |   | | ||| ||||| |||  ||||   ||| | ||   |   |  ||   |  |
Db    353 GCGCAATTTTATAATGAACTATCCAAACATATTAACATTAGTGTTTTTGGTGTGGTGTAC 294

Qy    933 GGAGAATATGTCAATGATAAAAATTTGATT---CCTACCATATCTGCTTGTAAATTTTAT 989
            ||          || |   ||  |  |   || ||| |  ||   ||||| |||||
Db    293 ACAGGGGTCAGGTTGGAGATTGATCAGTATTACCCGACCGTTGCTAGCTGTAAGTTTTAC 234

Qy    990 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTT 1049
          | || ||||| || ||| | ||||| |||||||||| || ||| |  ||    |  ||
Db    233 TTATCATTTGAGAACTCATTACACAAAGATTACATCACAGAGAAGGTCAACGNACCCCTT 174

Qy   1050 CTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATT 1109
            || ||  | || ||||| ||||||||||  | || |  ||||| || |  |   |
Db    173 GCAGCAGGAACCGTTCCTGTAGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTTC 114

Qy   1110 CCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTAT 1169
          ||  |||| || || ||||| || || |||| |     ||| |  |  || ||| || ||
Db    113 CCTTCAGACTCCTTTATTCACGTAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCAT 54

Qy   1170 CTGAAGGAAGTCGACAA-AAACAATAAGTTATACCTTAGTTACTTTAACTGGA 1221
          |||       | |||||   |  |  ||  |||| |  | |||||| | ||||
Db     53 CTGCTCTCCTTGGACAAGGNATGAGGAGGCATACATGCGCTACTTTGAGTGGA 1
```

Search completed: October 6, 2002, 07:16:35
Job time : 1134.97 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 27.4635 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-5_COPY_29_1374
Perfect score: 1346
Sequence: 1 actgctctcccctgcagctc.....agaaatggttttggaattaa 1346

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	9.4	1814	2	US-08-483-151-1	Sequence 1, Appli
2	127	9.4	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	113.4	8.4	1654	1	US-07-914-281-13	Sequence 13, Appl
4	113.4	8.4	1654	1	US-08-393-246-13	Sequence 13, Appl
5	113.4	8.4	1654	1	US-08-525-058A-13	Sequence 13, Appl
6	113.4	8.4	1654	2	US-08-696-731-13	Sequence 13, Appl
7	113.4	8.4	1654	4	US-09-042-531-13	Sequence 13, Appl
8	112.2	8.3	1086	1	US-07-914-281-12	Sequence 12, Appl
9	112.2	8.3	1086	1	US-08-393-246-12	Sequence 12, Appl
10	112.2	8.3	1086	1	US-08-525-058A-12	Sequence 12, Appl
11	112.2	8.3	1086	2	US-08-696-731-12	Sequence 12, Appl
12	112.2	8.3	1086	4	US-09-042-531-12	Sequence 12, Appl
13	112.2	8.3	2042	3	US-09-063-237-2	Sequence 2, Appli
14	112.2	8.3	2043	1	US-07-914-281-1	Sequence 1, Appli
15	112.2	8.3	2043	1	US-08-393-246-1	Sequence 1, Appli
16	112.2	8.3	2043	1	US-08-273-411-4	Sequence 4, Appli
17	112.2	8.3	2043	1	US-08-525-058A-1	Sequence 1, Appli
18	112.2	8.3	2043	2	US-08-696-731-1	Sequence 1, Appli
19	112.2	8.3	2043	4	US-09-042-531-1	Sequence 1, Appli
20	112.2	8.3	2043	5	PCT-US91-00899-1	Sequence 1, Appli
21	111.8	8.3	1316	1	US-07-914-281-10	Sequence 10, Appl
22	111.8	8.3	1316	1	US-08-393-246-10	Sequence 10, Appl
23	111.8	8.3	1316	1	US-08-525-058A-10	Sequence 10, Appl
24	111.8	8.3	1316	2	US-08-696-731-10	Sequence 10, Appl
25	111.8	8.3	1316	4	US-09-042-531-10	Sequence 10, Appl
26	100.8	7.5	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	100.8	7.5	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	100.8	7.5	1488	1	US-07-914-281-9	Sequence 9, Appli
29	100.8	7.5	1488	1	US-08-393-246-9	Sequence 9, Appli
30	100.8	7.5	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	100.8	7.5	1488	2	US-08-696-731-9	Sequence 9, Appli
32	100.8	7.5	1488	4	US-09-042-531-9	Sequence 9, Appli
33	100.8	7.5	2175	4	US-08-482-073-9	Sequence 9, Appli
34	100.8	7.5	2861	4	US-08-482-073-10	Sequence 10, Appl
35	100.8	7.5	3647	1	US-07-914-281-7	Sequence 7, Appli
36	100.8	7.5	3647	1	US-08-393-246-7	Sequence 7, Appli
37	100.8	7.5	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	100.8	7.5	3647	2	US-08-696-731-7	Sequence 7, Appli
39	100.8	7.5	3647	4	US-09-042-531-7	Sequence 7, Appli
40	100.8	7.5	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	97.6	7.3	2134	2	US-08-483-151-3	Sequence 3, Appli
42	97.6	7.3	2134	5	PCT-US96-06427-3	Sequence 3, Appli
43	47	3.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c 44	43.2	3.2	1776	1	US-08-722-001-29	Sequence 29, Appl
c 45	43.2	3.2	2140	1	US-08-334-698-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 9.4%; Score 127; DB 2; Length 1814;
Best Local Similarity 52.3%; Pred. No. 4.8e-27;
Matches 330; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

```
Qy    535 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 594
          | |||| ||| |   | ||||    |||     ||   |  |||| ||  |  |||| |
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy     595 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 654
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     655 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 714
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     715 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 774
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     775 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 834
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     835 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 894
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     895 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 954
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     955 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1014
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    1015 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1071
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    1072 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1131
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    1132 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

```
RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;    APPLICANT:  The General Hospital Corporation
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             9.4%;  Score 127;  DB 5;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 4.8e-27;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     535 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 594
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     595 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 654
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     655 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 714
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     715 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 774
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     775 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 834
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     835 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 894
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     895 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 954
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     955 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 1014
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy    1015 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 1071
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy    1072 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 1131
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy    1132 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 1162
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175
```

```
RESULT 3
US-07-914-281-13
; Sequence 13, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-13

  Query Match             8.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 4.1e-23;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
          || ||| ||||| |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
           |      ||     |  ||| |  ||||  | |||||     ||||      |  | |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755
          ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
          |||     ||| || ||||           | | |  |     |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
          |  |  |   |||      |  ||||| ||||| |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
          ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
          |    |    |    |  ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163
          |   | ||| | ||  | ||||| || ||||| || |   |    |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223
            ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1224 AAG 1226
            ||
Db   1095 GAG 1097


RESULT 4
US-08-393-246-13
; Sequence 13, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-13

  Query Match             8.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 4.1e-23;
```

```
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
           || ||| |||||  |||||| |||    |  || | |  ||  |   |||| || |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
             |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
             | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
           |||      ||| || ||||         | | | |     |  | |    | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
             |  |  |   |||      |  ||||| |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
           ||||| ||    || ||| |    ||            |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
             |       |      |   ||   ||||    ||  | ||     |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043
           |||||  |||| || || ||  | |||   || ||||||||  || |||||  |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy    1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103
           ||  |   |||  |  | || || || || ||||| ||     ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy    1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163
           |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy    1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223
             ||| ||| |||| | |||||  || |       |||||  || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    1224 AAG 1226
            ||
Db    1095 GAG 1097


RESULT 5
US-08-525-058A-13
; Sequence 13, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-13

  Query Match             8.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 4.1e-23;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy     462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
           || ||| |||||  |||||| |||    |  || | |  ||  |   |||| || |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy     522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
             |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
             | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
           |||      ||| || ||||         | | | |     |  | |    | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
             |  |  |   |||      |  ||||| |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
           ||||| ||    || ||| |    ||            |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
             |       |      |   ||   ||||    ||  | ||     |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043
           |||||  |||| || || ||  | |||   || ||||||||  || |||||  |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914
```

```
Qy   1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163
          |   | ||| | ||  | ||||| || ||||| || |   |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223
            ||| ||| ||||  | |||||  || |       |||||| || |||||| |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1224 AAG 1226
           ||
Db   1095 GAG 1097


RESULT 6
US-08-696-731-13
; Sequence 13, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-13

  Query Match             8.4%;  Score 113.4;  DB 2;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 4.1e-23;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
          || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
           |      ||      |  ||| |  |||| | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
          || ||  |  | || ||||||||  |||  |  | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
          | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755
          ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
          |||     ||| || ||||          | | | |      |    | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
          |  |  |   |||      |  ||||| ||||| |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
          ||||| ||    || ||| |    ||            |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
          |      |     |   ||  ||||   ||  | ||    |   || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043
          |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103
          ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163
          |   | ||| | ||  | ||||| || ||||| || |   |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223
            ||| ||| ||||  | |||||  || |       |||||| || |||||| |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1224 AAG 1226
           ||
Db   1095 GAG 1097


RESULT 7
US-09-042-531-13
; Sequence 13, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
```

```
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-042-531-13

  Query Match             8.4%;  Score 113.4;  DB 4;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 4.1e-23;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy    462 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 521
          || ||| |||||  ||||||  |||     |  || | |  ||  |   ||||  ||  |||
Db    318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy    522 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 578
          |    ||     |  ||| |  ||||  | |||||     ||||     |  |  |
Db    378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy    579 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 635
          || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db    438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy    636 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 695
          | ||||  |     |||  ||||| ||| | |   |||| || ||||  |||
Db    498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy    696 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 755
          ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db    558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy    756 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 806
          |||     ||| || ||||           | | | |       |     | |     | |
Db    618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy    807 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 866
          |  |  |   |||      |  |||||  |||||  |||   ||||||   ||  |
Db    678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy    867 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 926
          |||||  ||    ||  |||  |    ||           |  |  |  |    |||||   |
Db    738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy    927 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 986
          |      |      |      |   ||   ||||    ||  | ||    |  || ||
Db    795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy    987 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 1043
          |||||  |||| || || ||  | |||   || |||||||| || |||||  |     ||
Db    855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy   1044 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 1103
          ||  |   |||  |  | || || || || ||||| ||     ||    |||| ||||
Db    915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy   1104 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 1163
          |   | ||| | ||  | ||||| || ||||| || |   |    |||| || || ||
Db    975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy   1164 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 1223
            ||| ||| |||  | ||||  || |       ||||| || |||||  |||| ||
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy   1224 AAG 1226
            ||
Db   1095 GAG 1097


RESULT 8
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             8.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 7e-23;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;
```

```
Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
          | || ||||  |   |   |  | ||    | || |       | |||||  |      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
          |||||  ||| | ||  |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |   | |   |          | | |  | |   |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||           | | |  |    ||||| |   |                  |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          || |  | |  ||||    || | ||    |  || || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||     ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || |||||  || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |      ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 9
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             8.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 7e-23;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
          | || ||||  |   |   |  | ||    | || |       | |||||  |      |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
          |||||  ||| | ||  |||||  ||||  || ||   |  ||   |  | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |   | |   |          | | |  | |   |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||           | | |  |    ||||| |   |                  |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          || |  | |  ||||    || | ||    |  || || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||| || ||||| |     || ||  |  |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||     ||    ||||| ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || |||||  || |   |    ||||  || || ||   ||| ||| ||||  |
```

```
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |       |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 10
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             8.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 7e-23;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
          | || ||||  |   |   |  | ||    | || |      | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |    |  |    |     |          | |  |  |  |   |||
```

```
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
            |  |||||  |||||  |||   ||||||||  || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
            |    ||           |  |  |  |    |||||  |    |             |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |       |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 11
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
```

```
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match             8.3%;  Score 112.2;  DB 2;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 7e-23;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
            | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |    | |   |       |                | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||            |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||||| || ||||| |      || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 12
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match             8.3%;  Score 112.2;  DB 4;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 7e-23;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
            | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |    | |   |       |                | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||            |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690
```

```
Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| |||| |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |      ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 13
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             8.3%;  Score 112.2;  DB 3;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 1e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
          | || ||||  |   |    |  | ||    | || |       | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
            ||||| ||| | || ||||  ||||  || ||    |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |  ||  |   |  |   |  |          | | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
          |  |||||  ||||  |||   |||||||| || ||    ||||| ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||          |  |  |  |    ||||| |   |                |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| |||| |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |      ||||| || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044


RESULT 14
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             8.3%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 1e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
          | || ||||  |   |   | || |     | || |       | |||||  |      |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
          |   ||  |   |  |     |     |           | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944
          |    ||          |  |  |  |    |||||  |   |              |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004
          ||  |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy   1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy   1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy   1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181
          ||||| || ||||| || |    |    ||||  || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy   1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226
          |||||  || |       ||||| || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 15
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             8.3%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 1e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    537 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 596
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    597 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 653
          | || ||||  |   |   | || |     | || |       | |||||  |      |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    654 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 713
             ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    714 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 773
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    774 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 824
```

```
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     825 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 884

Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     885 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 944

Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     945 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 1004

Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    1005 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 1061

Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    1062 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 1121

Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    1122 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 1181

Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    1182 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 1226

Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:22:15
Job time : 39.4635 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:39:45 ; Search time 1146.59 Seconds
(without alignments)
16699.717 Million cell updates/sec

Title: US-09-744-748-5_COPY_460_1374
Perfect score: 915
Sequence: 1 accaaaactgattattttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	915	100.0	3019	9	AB023021	AB023021 Homo sapi
	2	913.4	99.8	80247	9	AL512406	AL512406 Human DNA
	3	910.2	99.5	2501	9	HSA238701	AJ238701 Homo sapi
	4	792	86.6	1128	10	AF345993	AF345993 Rattus no
	5	791.8	86.5	2139	10	AB015426	AB015426 Mus muscu
	6	790.4	86.4	2156	10	AB049819	AB049819 Rattus no
	7	783.8	85.7	1705	10	AF230460	AF230460 Cricetulu
	8	666	72.8	1707	5	AB035906	AB035906 Gallus ga
	9	581.2	63.5	1080	5	AB035905	AB035905 Xenopus l
	10	580.2	63.4	71396	2	AC100674	AC100674 Mus muscu
	11	277.8	30.4	1567	5	AB023627	AB023627 Danio rer
	12	234.6	25.6	1832	5	AB023628	AB023628 Danio rer
	13	222.6	24.3	1092	5	AB035907	AB035907 Xenopus l
	14	220.2	24.1	1264	5	AF288369	AF288369 Gallus ga
	15	176.2	19.3	2657	5	GGU73678	U73678 Gallus gall
	16	145.2	15.9	1429	10	CGU78737	U78737 Cricetulus
	17	145.2	15.9	1462	10	AF090450	AF090450 Cricetulu
	18	143.6	15.7	1718	10	AF090449	AF090449 Cricetulu
	19	137.2	15.0	1044	5	AB035908	AB035908 Gallus ga
	20	132	14.4	2293	4	BTA134FTG	X87810 B.taurus al
	21	132	14.4	12607	4	BTA132772	AJ132772 Bos tauru
c	22	128	14.0	71396	2	AC100674	AC100674 Mus muscu
	23	127.6	13.9	1158	10	AB039126	AB039126 Mus muscu
	24	127	13.9	1487	3	AF016899	AF016899 Schistoso
	25	127	13.9	1814	6	AR028686	AR028686 Sequence
	26	127	13.9	3594	10	MMU45980	U45980 Mus musculu
c	27	127	13.9	117993	2	AC091284	AC091284 Mus muscu
	28	125.6	13.7	1613	4	BTA132776	AJ132776 Bos tauru
	29	125.6	13.7	1619	4	BTA132775	AJ132775 Bos tauru
	30	125.6	13.7	1644	4	BTA132774	AJ132774 Bos tauru
	31	125.6	13.7	1873	4	BTA132773	AJ132773 Bos tauru
	32	124.6	13.6	1119	9	AF345881	AF345881 Macaca mu
	33	122.8	13.4	1170	10	AB039127	AB039127 Mus muscu
	34	122.8	13.4	1170	10	AB039131	AB039131 Mus muscu
	35	122.8	13.4	1615	10	MMU33457	U33457 Mus musculu
	36	122	13.3	1113	9	AF345884	AF345884 Saimiri s
	37	121.2	13.2	1170	10	AB039124	AB039124 Mus muscu
	38	121.2	13.2	1170	10	AB039125	AB039125 Mus muscu
	39	121.2	13.2	1170	10	AB039128	AB039128 Mus muscu
	40	121.2	13.2	1170	10	AB039129	AB039129 Mus muscu
	41	121.2	13.2	1170	10	AB039130	AB039130 Mus muscu
	42	121.2	13.2	1170	10	AB039132	AB039132 Mus muscu
	43	119.8	13.1	1119	9	AF345883	AF345883 Macaca mu
	44	118.4	12.9	1170	10	AB039133	AB039133 Mus spici
	45	118.2	12.9	1080	9	AB051859	AB051859 Pongo pyg

ALIGNMENTS

RESULT 1
AB023021
LOCUS AB023021 3019 bp mRNA linear PRI 23-JUN-1999
DEFINITION Homo sapiens FUT9 mRNA for alpha-1,3-fucosyltransferase IX, complete cds.
ACCESSION AB023021
VERSION AB023021.1 GI:5139692
KEYWORDS alpha-1,3-fucosyltransferase IX; FUT9.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kaneko,M., Kudo,T., Iwasaki,H., Ikehara,Y., Nishihara,S., Nakagawa,S., Sasaki,K., Shiina,T., Inoko,H., Saitou,N. and Narimatsu,H.
TITLE Alpha 1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX
JOURNAL FEBS Lett. 453, 237-242 (1999)
REFERENCE 2 (bases 1 to 3019)
AUTHORS Kaneko,M., Kudo,T. and Narimatsu,H.
TITLE Direct Submission

```
  JOURNAL   Submitted (29-JAN-1999) Mika Kaneko, Institute of Life Science,
            Soka University, Division of Cell Biology; Soka University, 1-236
            Tangi-cho, Hachioji, Tokyo 192-8577, Japan
            (E-mail:mika@t.soka.ac.jp, Tel:81-426-91-2495(ex.5132),
            Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .3019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            295. .1374
                     /gene="FUT9"
     CDS             295. .1374
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase IX"
                     /protein_id="BAA81685.1"
                     /db_xref="GI:5139693"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      937 a    594 c    541 g    947 t
ORIGIN

  Query Match             100.0%;  Score 915;  DB 9;  Length 3019;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-231;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 519

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 579

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 639

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 699

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 759

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 819

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 879

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 939

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 999

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 1059

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1060 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 1119

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 1179

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1180 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1239

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1240 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1299

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1300 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1359

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db    1360 TGGTTTTGGAATTAA 1374


RESULT 2
AL512406
LOCUS       AL512406               80247 bp    DNA     linear   PRI 27-APR-2001
DEFINITION  Human DNA sequence from clone RP11-504J9 on chromosome 6, complete
            sequence.
ACCESSION   AL512406
VERSION     AL512406.14  GI:13897154
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80247)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2001 this sequence version replaced gi:13396709.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-504J9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-504J9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-381A2 is at 80148 in this sequence.
            The true right end of clone RP11-77B15 is at 100 in this sequence.
```

```
FEATURES             Location/Qualifiers
     source          1. .80247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-504J9"
                     /clone_lib="RPCI-11.2"
     repeat_region   184. .1422
                     /note="L2 repeat: matches 875. .2419 of consensus"
     repeat_region   1744. .1834
                     /note="MIR repeat: matches 166. .258 of consensus"
     repeat_region   2950. .2977
                     /note="14 copies 2 mer ca 100% conserved"
     repeat_region   3411. .3669
                     /note="LTR16C repeat: matches 93. .364 of consensus"
     repeat_region   4372. .4704
                     /note="MLT1A1 repeat: matches 5. .365 of consensus"
     misc_feature    complement(7821. .8406)
                     /note="match: GSS: Em:AQ550748"
     repeat_region   8126. .8184
                     /note="MER58B repeat: matches 272. .332 of consensus"
     repeat_region   8185. .8696
                     /note="MLT2D repeat: matches 1. .513 of consensus"
     repeat_region   8779. .8817
                     /note="MLT2D repeat: matches 504. .542 of consensus"
     repeat_region   8819. .8974
                     /note="MER58A repeat: matches 6. .162 of consensus"
     misc_feature    complement(9443. .9891)
                     /note="match: GSS: Em:B90633"
     misc_feature    complement(9452. .9883)
                     /note="match: GSS: Em:AQ119695"
     repeat_region   9782. .9894
                     /note="MIR repeat: matches 92. .205 of consensus"
     repeat_region   10089. .10212
                     /note="MIR repeat: matches 135. .262 of consensus"
     misc_feature    complement(10093. .10671)
                     /note="match: GSS: Em:AQ237061"
     repeat_region   10223. .10282
                     /note="MIR repeat: matches 203. .262 of consensus"
     misc_feature    complement(10315. .10631)
                     /note="match: GSS: Em:AQ200259"
     misc_feature    10805. .11186
                     /note="match: GSS: Em:AQ072353"
     repeat_region   11657. .11844
                     /note="L2 repeat: matches 2546. .2750 of consensus"
     misc_feature    complement(11659. .12084)
                     /note="match: GSS: Em:AQ518361"
     misc_feature    complement(11674. .11987)
                     /note="match: GSS: Em:AQ094678"
     misc_feature    12083. .12588
                     /note="match: GSS: Em:AQ691235"
     repeat_region   12596. .12810
                     /note="MIR repeat: matches 23. .248 of consensus"
     repeat_region   12920. .12986
                     /note="L1MA5 repeat: matches 6228. .6295 of consensus"
     repeat_region   12987. .13169
                     /note="L1MA5 repeat: matches 5852. .6043 of consensus"
     repeat_region   13168. .13280
                     /note="L1MA8 repeat: matches 5710. .5832 of consensus"
     repeat_region   13851. .17218
                     /note="L1PB1 repeat: matches 2787. .6155 of consensus"
     repeat_region   18903. .18953
                     /note="MLT2CB repeat: matches 439. .501 of consensus"
     repeat_region   18984. .19433
                     /note="MLT2CB repeat: matches 1. .460 of consensus"
     misc_feature    complement(19108. .19690)
                     /note="match: GSS: Em:AQ563935"
     repeat_region   20683. .20801
                     /note="AluJo/FRAM repeat: matches 169. .288 of consensus"
     repeat_region   20802. .20862
                     /note="AluJ/FLAM repeat: matches 3. .66 of consensus"
     repeat_region   20912. .21033
                     /note="MIR repeat: matches 114. .237 of consensus"
     repeat_region   21047. .21128
                     /note="41 copies 2 mer ta 63% conserved"
     repeat_region   21676. .21859
                     /note="L2 repeat: matches 2130. .2308 of consensus"
     repeat_region   21890. .23612
                     /note="L1P repeat: matches 841. .2570 of consensus"
     repeat_region   23601. .26609
                     /note="L1PA5 repeat: matches 3107. .6145 of consensus"
     repeat_region   26620. .26789
                     /note="MER5A repeat: matches 1. .174 of consensus"
     repeat_region   28521. .28627
                     /note="MER39 repeat: matches 5. .111 of consensus"
     repeat_region   28656. .28837
                     /note="MER39b repeat: matches 54. .235 of consensus"
     repeat_region   28838. .29142
                     /note="AluSc repeat: matches 4. .308 of consensus"
     repeat_region   29143. .29494
                     /note="MER39b repeat: matches 235. .549 of consensus"
     repeat_region   30529. .31344
                     /note="L1PA3 repeat: matches 5331. .6146 of consensus"
     repeat_region   32234. .32340
                     /note="L2 repeat: matches 2627. .2750 of consensus"
     repeat_region   32399. .32578
                     /note="MIR repeat: matches 12. .203 of consensus"
     repeat_region   32816. .32979
                     /note="MIR repeat: matches 91. .259 of consensus"
     misc_feature    complement(33006. .33450)
                     /note="match: GSS: Em:AQ261227"
     misc_feature    complement(33058. .33444)
                     /note="match: GSS: Em:AQ040453"
     repeat_region   33312. .33596
                     /note="L1M4 repeat: matches 3473. .3773 of consensus"
     repeat_region   34856. .34915
                     /note="30 copies 2 mer aa 68% conserved"
     misc_feature    complement(34987. .35547)
                     /note="match: GSS: Em:AQ321598
                     match: STS: Em:G55571"
     misc_feature    complement(36488. .36936)
                     /note="match: GSS: Em:B41073"
     misc_feature    complement(36654. .37090)
                     /note="match: GSS: Em:AQ772584"
     misc_feature    37102. .37591
                     /note="match: GSS: Em:AQ489567"
     misc_feature    complement(37581. .38221)
                     /note="match: GSS: Em:AQ418233"
     repeat_region   37720. .37772
                     /note="L2 repeat: matches 2657. .2710 of consensus"
     misc_feature    complement(37778. .38215)
                     /note="match: GSS: Em:AQ819773"
     misc_feature    38218. .38732
                     /note="match: GSS: Em:AQ599868"
     repeat_region   38861. .39164
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   39612. .39653
                     /note="21 copies 2 mer ac 90% conserved"
     repeat_region   40854. .40909
                     /note="L2 repeat: matches 2655. .2710 of consensus"
     repeat_region   42169. .42270
                     /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   42411. .42548
                     /note="MER5B repeat: matches 21. .181 of consensus"
     repeat_region   42780. .42938
                     /note="MER5B repeat: matches 17. .177 of consensus"
     repeat_region   43229. .43326
                     /note="MIR repeat: matches 40. .148 of consensus"
     repeat_region   45383. .45428
                     /note="MER5B repeat: matches 18. .62 of consensus"
     misc_feature    complement(47284. .47971)
                     /note="match: STS: Em:G51995
                     match: GSS: Em:AQ269059"
     misc_feature    complement(47537. .48011)
                     /note="match: STS: Em:G36899"
     misc_feature    complement(47723. .48014)
```

```
                  /note="match: STS: Em:G30049"
     misc_feature    47972. .48451
                  /note="match: GSS: Em:AQ336527"
     misc_feature    complement(48518. .48985)
                  /note="match: GSS: Em:AQ801834"
     repeat_region   48597. .48803

  Query Match             99.8%;  Score 913.4;  DB 9;  Length 80247;
  Best Local Similarity   99.9%;  Pred. No. 2.2e-230;
  Matches  914;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35724 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 35783

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35784 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 35843

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35844 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 35903

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35904 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 35963

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35964 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 36023

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36024 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 36083

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36084 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 36143

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36144 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 36203

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36204 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 36263

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36264 TCTACTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 36323

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36324 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 36383

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36384 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 36443

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36444 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 36503

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36504 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 36563

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36564 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 36623

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db   36624 TGGTTTTGGAATTAA 36638


RESULT 3
HSA238701
LOCUS       HSA238701               2501 bp    mRNA    linear   PRI 21-OCT-2000
DEFINITION  Homo sapiens mRNA for alpha-3-fucosyltransferase.
ACCESSION   AJ238701
VERSION     AJ238701.1  GI:4741566
KEYWORDS    alpha-3-fucosyltransferase; FUT9 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Cailleau-Thomas,A., Coullin,P., Candelier,J.J., Balanzino,L.,
            Mennesson,B., Oriol,R. and Mollicone,R.
  TITLE     FUT4 and FUT9 genes are expressed early in human embryogenesis
  JOURNAL   Glycobiology 10 (8), 789-802 (2000)
  MEDLINE   20386991
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Mollicone,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Mollicone R., Glycobiology, INSERM U504, 16
            Av. Paul Vaillant-Couturier, Villejuif, 94807 Cedex, FRANCE
COMMENT     related sequence:  AB015426.
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /tissue_type="whole embryo"
                     /dev_stage="40 to 60 days"
     gene            105. .2466
                     /gene="FUT9"
     CDS             105. .1184
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha-3-fucosyltransferase"
                     /protein_id="CAB41890.1"
                     /db_xref="GI:4741567"
                     /db_xref="SPTREMBL:Q9Y231"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISACKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2192
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2235
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2310
                     /gene="FUT9"
                     /evidence=experimental
     polyA_site      2466
                     /gene="FUT9"
                     /evidence=experimental
BASE COUNT      818 a    436 c    422 g    825 t
ORIGIN

  Query Match             99.5%;  Score 910.2;  DB 9;  Length 2501;
  Best Local Similarity   99.7%;  Pred. No. 1.6e-229;
  Matches  912;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 329

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
```

```
Db    330 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 389

Qy    121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    390 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGCTGG 449

Qy    181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          |||||||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db    450 GATCTGACTAACTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 509

Qy    241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    510 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 569

Qy    301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    570 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 629

Qy    361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    630 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 689

Qy    421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    690 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 749

Qy    481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    750 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 809

Qy    541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    810 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 869

Qy    601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    870 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 929

Qy    661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 989

Qy    721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    990 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1049

Qy    781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1050 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1109

Qy    841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1169

Qy    901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db   1170 TGGTTTTGGAATTAA 1184


RESULT 4
AF345993
LOCUS       AF345993               1128 bp    mRNA    linear   ROD 07-MAR-2001
DEFINITION  Rattus norvegicus alpha1,3-fucosyltransferase IX (Fut9) mRNA,
            complete cds.
ACCESSION   AF345993
VERSION     AF345993.1  GI:13242183
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Baboval,T., Henion,T., Kinnally,E. and Smith,F.I.
  TITLE     Molecular cloning of rat alpha1,3-fucosyltransferase IX (Fuc-TIX)
            and comparison of the expression of fuc-TIV and fuc-TIX genes
            during rat postnatal cerebellum development
  JOURNAL   J. Neurosci. Res. 62 (2), 206-215 (2000)
   PUBMED   11020213
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Smith,F.I. and Baboval,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Biomedical Sciences, Eunice Kennedy Shriver
            Center, 200 Trapelo Rd., Waltham, MA, USA
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1128
                     /gene="Fut9"
     CDS             49. .1128
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="AAK16591.1"
                     /db_xref="GI:13242184"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      324 a    256 c    228 g    320 t
ORIGIN

  Query Match           86.6%;  Score 792;  DB 10;  Length 1128;
  Best Local Similarity   91.8%;  Pred. No. 2.5e-198;
  Matches  837;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy      4 AAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACC 63
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    217 AAAACTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACC 276

Qy     64 TTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGAC 123
          |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db    277 TTTGACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGAC 336

Qy    124 CGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGAT 183
          || ||  |||||||||||||||||| || ||||| || ||| |||||||||| ||||||
Db    337 CGCTCGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGAT 396

Qy    184 CTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTG 243
          ||||| || |||||||||| || ||||| ||||| ||||||||||||||||||||||| ||
Db    397 CTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTA 456

Qy    244 GAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTG 303
          || ||||| |||||||| |||||||||||||||||||||||||||| ||||||||||
Db    457 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTA 516

Qy    304 ACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAAT 363
          ||||| |||||||||||||| |||||||||| |||||||||||||| |||||||
Db    517 ACTTATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGC 576

Qy    364 CCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGG 423
          ||||| |||||||||||||||||||||||||| || ||||||||||||||||||||||
Db    577 CCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGG 636

Qy    424 AACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCAT 483
          || || |||||||||| |||||||||||||| ||||| ||||| || ||||||||||
Db    637 AATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 696

Qy    484 ACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCT 543
          ||||| |||||||||||||||||||||| |||||||||| ||||||| |||||||||
```

```
Db     697 ACCTATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCT 756

Qy     544 GCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAG 603
           |||||||||||||||||||||||||||||||| ||||| ||||| |||||||||||| ||||||
Db     757 ACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 816

Qy     604 CTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAAC 663
           || |||||||| ||| ||||||| |||||||||||||| ||||| ||||||||||||||||
Db     817 CTCTACAATGCATTTTTGGCTGGTTCTGTACCTGTTGTCCTGGGTCCATCTAGGGAAAAC 876

Qy     664 TATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGT 723
           |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db     877 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCCAGT 936

Qy     724 GAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTT 783
           ||| | ||||| |||||||||| || ||||||||||||||||||| ||||||||||
Db     937 GAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTC 996

Qy     784 AACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTG 843
           |||||||||||||||| |||||||| || ||||| ||||||||||||||||||||  ||
Db     997 AACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTG 1056

Qy     844 GCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGG 903
           || |||||||||| |||||||||||||||||||||||||| ||||||||||||
Db    1057 GCATGCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGG 1116

Qy     904 TTTTGGAATTAA 915
           ||||||||||||
Db    1117 TTTTGGAATTAA 1128
```

```
RESULT 5
AB015426
LOCUS       AB015426               2139 bp    mRNA    linear   ROD 23-JUN-1999
DEFINITION  Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, complete
            cds.
ACCESSION   AB015426
VERSION     AB015426.1  GI:3702718
KEYWORDS    Fut9; alpha1,3-fucosyltransferase IX.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kudo,T., Ikehara,Y., Togayachi,A., Kaneko,M., Hiraga,T., Sasaki,K.
            and Narimatsu,H.
  TITLE     Expression cloning and characterization of a novel murine alpha1,
            3-fucosyltransferase, mFuc-TIX, that synthesizes the Lewis x (CD15)
            epitope in brain and kidney
  JOURNAL   J. Biol. Chem. 273 (41), 26729-26738 (1998)
  MEDLINE   98434588
REFERENCE   2  (bases 1 to 2139)
  AUTHORS   Kudo,T. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1998) Takashi Kudo, Institute of Life Science,
            Soka University, Division of Cell Biology; 1-236 Tangi-cho,
            Hachioji, Tokyo 192-8577, Japan (E-mail:tkudo@t.soka.ac.jp,
            Tel:+81-426-91-2495, Fax:+81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2139
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            115. .1194
                     /gene="Fut9"
     CDS             115. .1194
                     /gene="Fut9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAA33522.1"
                     /db_xref="GI:3702719"
                     /translation="MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
     polyA_site      2139
                     /note="31 a nucleotides"
BASE COUNT      684 a    395 c    368 g    692 t
ORIGIN
```

```
  Query Match           86.5%;  Score 791.8;  DB 10;  Length 2139;
  Best Local Similarity  91.6%;  Pred. No. 2.8e-198;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           || |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db     280 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 339

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db     340 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 399

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db     400 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 459

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db     460 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 519

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| ||||||||
Db     520 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 579

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     580 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 639

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db     640 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 699

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           |||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db     700 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 759

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db     760 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 819

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||||| |||
Db     820 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 879

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db     880 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 939

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           |||||||||||||||||||||||| ||||||||||||||||||||||||||| ||||||
Db     940 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 999

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db    1000 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 1059

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           |||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db    1060 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1119

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| |||||| |||||||||||||||||||||||||||||||||||| |||
Db    1120 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1179
```

```
Qy    901 TGGTTTTGGAATTAA 915
          ||||||||||||||||
Db   1180 TGGTTTTGGAATTAA 1194


RESULT 6
AB049819
LOCUS       AB049819               2156 bp    mRNA    linear   ROD 18-JAN-2002
DEFINITION  Rattus norvegicus FUT9 mRNA for alpha1,3-fucosyltransferase IX,
            complete cds.
ACCESSION   AB049819
VERSION     AB049819.1  GI:13591588
KEYWORDS    .
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Shimoda,Y., Tajima,Y., Osanai,T., Katsume,A., Kohara,M., Kudo,T.,
            Narimatsu,H., Takashima,N., Ishii,Y., Nakamura,S., Osumi,N. and
            Sanai,Y.
  TITLE     Pax6 Controls the Expression of Lewis x Epitope in the Embryonic
            Forebrain by Regulating alpha 1,3-Fucosyltransferase IX Expression
  JOURNAL   J. Biol. Chem. 277 (3), 2033-2039 (2002)
   PUBMED   11675393
REFERENCE   2  (bases 1 to 2156)
  AUTHORS   Sanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Yutaka Sanai, Tokyo Metropolitan Institute
            of Medical Science, Department of Biochemical Cell Research;
            Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113-8613, Japan
            (E-mail:sanai@rinshoken.or.jp, Tel:81-3-3823-2101(ex.5233),
            Fax:81-3-3828-6663)
FEATURES             Location/Qualifiers
     source          1. .2156
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            346. .1425
                     /gene="FUT9"
     CDS             346. .1425
                     /gene="FUT9"
                     /codon_start=1
                     /product="alpha1,3-fucosyltransferase IX"
                     /protein_id="BAB40953.1"
                     /db_xref="GI:13591589"
                     /translation="MTSTSKGILRPFLIVCIILGCFVACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSRKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTSPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      605 a    470 c    437 g    644 t
ORIGIN

  Query Match           86.4%;  Score 790.4;  DB 10;  Length 2156;
  Best Local Similarity   91.7%;  Pred. No. 6.5e-198;
  Matches  836;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      4 AAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACC 63
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    514 AAAACTGATTATTTTAATGAAACTACTATTCTGGTTTGGGTGTGGCCATTTGGGCAGACC 573

Qy     64 TTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGAC 123
          |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db    574 TTTGACCTTACATCCTGCCAAGCCATGTTCAACATCCAAGGGTGCCATCTCACAACGGAC 633

Qy    124 CGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGAT 183
          || ||  |||||||||||||||||| || ||||| || ||| |||||||||| ||||||
Db    634 CGCTCGTTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGGGAT 693

Qy    184 CTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTG 243
          ||||| || |||||||||| || ||||| ||||| ||||||||||||||||||||| ||
Db    694 CTGACTAACTTACCTCAGCAGGCCAGGCCGCCCTTTCAGAAATGGATTTGGATGAACTTA 753

Qy    244 GAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTG 303
          || ||||| |||||||| |||||||||||||||||||||||||||||| ||||||||||
Db    754 GAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAGCACTTGTTCAACCTGACTCTA 813

Qy    304 ACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAAT 363
          ||||| |||||||||||||||| |||||||||| |||||||||||||| |||||||
Db    814 ACTTATCGCCGTGATTCAGATATACAAGTGCCTTACGGCTTCTTGACGGTGAGCACAAGC 873

Qy    364 CCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGG 423
          ||||| |||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db    874 CCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGTTAGTGTGCTGGGTTGTGAGTAACTGG 933

Qy    424 AACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCAT 483
          || || |||||||||||| |||||||||||||| ||||| ||||| || ||||||||||
Db    934 AATCCCGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATCCAC 993

Qy    484 ACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCT 543
          ||||| |||||||||||||||||||||| |||||||||||| ||||||| |||||||||
Db    994 ACCTATGGGCAAGCATTTGGAGAATATGTGAATGATAAAAATCTGATTCCCACCATATCT 1053

Qy    544 GCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAG 603
          |||||||||||||||||||||||||||||| ||||| ||||| |||||||||| ||||||
Db   1054 ACTTGTAAATTTTATCTTTCCTTTGAAAACTCAATTCACAAAGATTACATCACAGAAAAG 1113

Qy    604 CTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAAC 663
          || |||||||| ||| ||||||| || |||||||||| ||||||||||||||||||||
Db   1114 CTCTACAATGCATTTTTGGCTGGTTCGGTACCTGTTGTCCTGGGACCATCTAGGGAAAAC 1173

Qy    664 TATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGT 723
          |||||||||||||||||||| |||||||||||||||||||||| |||||| ||||||
Db   1174 TATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCACCCAGT 1233

Qy    724 GAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTT 783
          ||| | ||||| |||||||||| || ||||||||||||||||||| ||||||||||
Db   1234 GAGTTGGCAAAATATCTGAAGGAGGTTGACAAAAACAATAAGTTATATCTTAGTTACTTC 1293

Qy    784 AACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTG 843
          |||||||||||||||| |||||||| || ||||| |||||||||||||||||||  ||
Db   1294 AACTGGAGGAAGGATTTTACTGTAAACCTCCCACGGTTTTGGGAATCACATGCATGCCTG 1353

Qy    844 GCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGG 903
          || |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db   1354 GCATGCGATCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGCAATTTAGAGAAATGG 1413

Qy    904 TTTTGGAATTAA 915
          ||||||||||||
Db   1414 TTTTGGAATTAA 1425


RESULT 7
AF230460
LOCUS       AF230460               1705 bp    mRNA    linear   ROD 12-JUL-2000
DEFINITION  Cricetulus griseus alpha(1,3)fucosyltransferase type IX (Fut9)
            mRNA, complete cds.
ACCESSION   AF230460
VERSION     AF230460.1  GI:9049663
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Patnaik,S.K., Zhang,A., Shi,S. and Stanley,P.
  TITLE     Alpha(1,3)fucosyltransferases expressed by the gain-of-function
            Chinese hamster ovary glycosylation mutants LEC12, LEC29, and LEC30
  JOURNAL   Arch. Biochem. Biophys. 375 (2), 322-332 (2000)
  MEDLINE   20166953
   PUBMED   10700388
REFERENCE   2  (bases 1 to 1705)
```

```
  AUTHORS   Patnaik,S.K., Shi,S. and Stanley,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) Cell Biology, Albert Einstein College of
            Medicine, Chanin 516, 1300 Morris Park Avenue, New York, NY 10461,
            USA
FEATURES             Location/Qualifiers
     source          1. .1705
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
     gene            1. .1705
                     /gene="Fut9"
                     /note="cgFut9"
     CDS             62. .1141
                     /gene="Fut9"
                     /note="Fuc-TIX"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase type IX"
                     /protein_id="AAF82412.1"
                     /db_xref="GI:9049664"
                     /translation="MTSTSKGILRPFLIVCVILACFMACLLIYIKPTNSWVFSPMESA
                     SSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNK
                     SHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTY
                     GQAFGEYVNEKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      515 a    336 c    320 g    534 t
ORIGIN

  Query Match             85.7%;  Score 783.8;  DB 10;  Length 1705;
  Best Local Similarity   91.0%;  Pred. No. 3.6e-196;
  Matches  833;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          || |||||||||||||||||||||| |||||||| || ||||||||||||||||||||||
Db     227 ACAAAAACTGATTATTTTAATGAAACCACTATTCTTGTTTGGGTGTGGCCATTTGGGCAG 286

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           |||||||||||||||||||||||| ||||||||||||||| |||||||||| ||
Db     287 ACCTTTGACCTTACATCCTGCCAAGCGATGTTCAACATCCAAGGGTGCCATCTCACTACA 346

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||| |||||||||||||||||||| || ||||| || ||| |||||||||| |||
Db     347 GACCGCTCACTGTACAACAAATCCCATGCGGTCCTGATACACCACAGAGACATCAGCTGG 406

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| |||||||| ||||| ||||||||||||||||||||
Db     407 GATCTGACTAACTTACCTCAGCAGGCTAGGCCGCCCTTTCAGAAATGGATTTGGATGAAC 466

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           | || || || ||||| || |||||||||||||||||||||| ||||| |||||||||
Db     467 CTAGAGTCTCCCACTCATACCCCCCAAAAGAGTGGCATTGAGCATTTGTTCAACCTGACT 526

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           || || || ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     527 CTAACATATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 586

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||||||  ||||||||||||||||||||||||
Db     587 AATCCCTTTGTGTTTGAAGTGCCAAGCAAAGAGAAGCTGGTGTGCTGGGTTGTGAGTAAC 646

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           ||||| || |||||||||| |||||||||| |||||||| ||||| || |||||||||
Db     647 TGGAATCCGGAGCATGCCAGGGTCAAGTATTATAATGAGCTCAGCAAGAGTATTGAAATC 706

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           |||||||| |||||||| |||||||||| || ||||| |||||| | ||||| ||||||
Db     707 CATACCTATGGGCAAGCCTTTGGAGAATACGTGAATGAAAAAAATCTAATTCCCACCATA 766

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db     767 TCTACTTGTAAATTTTATCTTTCTTTTGAAAACTCAATTCACAAAGATTACATCACCGAA 826

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||| |||||||| ||| |||||||||| |||||||||||||||||||||| ||||||
Db     827 AAGCTCTACAATGCATTTTTGGCTGGCTCAGTACCTGTTGTTCTGGGACCATCCAGGGAA 886

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     887 AACTATGAGAATTATATTCCAGCGGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 946

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           |||||| | ||||| ||||| |||||||| |||||||||||||||| ||||||||||
Db     947 AGTGAGTTGGCAAAATATCTTAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAT 1006

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           || |||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db    1007 TTCAACTGGAGGAAGGATTTCACTGTAAACCTCCCACGATTTTGGGAATCACATGCATGC 1066

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           ||||||| ||||| || |||||||| ||||| |||||||||| ||||||||||||||||
Db    1067 CTGGCTTGTGATCACGTAAAAAGGCACCAAGAGTATAAGTCTGTGGGTAATTTAGAGAAA 1126

Qy     901 TGGTTTTGGAATTAA 915
           |||||||||||||||
Db    1127 TGGTTTTGGAATTAA 1141


RESULT 8
AB035906
LOCUS       AB035906               1707 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Gallus gallus gene for cFUT9, complete cds.
ACCESSION   AB035906
VERSION     AB035906.1  GI:18146865
KEYWORDS    cFUT9.
SOURCE      Gallus gallus brain DNA.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
     gene            268. .1347
                     /gene="cFuc-TIX"
     CDS             268. .1347
                     /gene="cFuc-TIX"
                     /codon_start=1
                     /product="cFUT9"
                     /protein_id="BAB82489.1"
                     /db_xref="GI:18146866"
                     /translation="MTSTSKGIFRPFFIVFIVLGCFMAFILIYIKPTSSWISNPVESA
                     SSVLKMKSFFSSKTDNFNETIILVWLWPFGQTFDLTSCQTMFNIHGCHLTIDRSLYNK
                     SHAVLIHHRDINWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGSMIVGTGAFTYEVPSKENLVCWVVSNWNPEHARVKYYNELSKYIEIHTY
                     GQAFGDYVNDKNLIPTISSCKFYLSFENSIHKDYITEKLYNALLAGSVPVVLGPSREN
                     YENYIPADSFIHVEDFLSPRELAEYLLMLDKNNKMYLSYFNWRKDFSVHLPKFWESHA
                     CLACDHVKRHQEYKSIGNLEKWFWN"
BASE COUNT      517 a    338 c    329 g    521 t      2 others
ORIGIN

  Query Match             72.8%;  Score 666;  DB 5;  Length 1707;
```

```
  Best Local Similarity   83.0%;  Pred. No. 4.3e-165;
  Matches  759;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy       2 CCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGA 61
           |||||||||||| ||||||||||||||| |||| |||| ||| | ||||||||||| ||||
Db     434 CCAAAACTGATAATTTTAATGAAACTATTATTTTGGTTTGGCTTTGGCCATTTGGCCAGA 493

Qy      62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGG 121
           | || || || |||||||||||| | |||||||||||||| || |||||||| || |  |
Db     494 CATTCGATCTAACATCCTGCCAAACGATGTTCAACATCCATGGGTGCCATCTGACTATTG 553

Qy     122 ACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGG 181
           |||| ||||| || ||||| |||||||||||||| || |||||| | ||||| |  ||||
Db     554 ACCGCTCACTATATAACAAGTCCCATGCAGTTCTCATACATCACAGGGACATTAACTGGG 613

Qy     182 ATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATT 241
           ||||||| |||||||||||||||| |||||||| |||||||||||||||||||||||| |
Db     614 ATCTGACTAATTTACCTCAGCAAGCCAGGCCACCATTCCAGAAATGGATTTGGATGAACT 673

Qy     242 TGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTC 301
           ||||||| |||||||| ||||| |||||||||||||||| ||| | |||||||| || |
Db     674 TGGAATCTCCAACTCATACTCCACAAAAGAGTGGCATTGAACACCTCTTTAACCTAACCC 733

Qy     302 TGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAA 361
           |||||||||| |||||||||||||| || |||||||||||| | |||  ||  | |||
Db     734 TGACTTACCGGCGTGATTCAGATATTCAGGTGCCTTATGGCTCCATGATAGTTGGTACAG 793

Qy     362 ATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 421
              ||||    | |||||||||||| ||||| || ||||| || |||||||| |||||||
Db     794 GCGCCTTTACATATGAAGTGCCAAGTAAAGAAAATTTGGTTTGTTGGGTTGTAAGTAACT 853

Qy     422 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 481
           ||||||||||||||||  |||||||||||||||||||||| |||||   ||||||||||
Db     854 GGAACCCTGAGCATGCTCGAGTCAAGTATTACAATGAGCTTAGCAAGTACATTGAAATCC 913

Qy     482 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 541
           |||||||||||||||| |||||||| || ||||| |||||||| || ||||| || || |
Db     914 ATACCTACGGGCAAGCCTTTGGAGACTACGTCAACGATAAAAACTTAATTCCAACTATCT 973

Qy     542 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 601
           |   ||| ||||| || ||||| |||||||||||||||||||| |||||||| || || |
Db     974 CCAGTTGCAAATTCTACCTTTCATTTGAAAATTCAATCCACAAAGATTACATTACTGAGA 1033

Qy     602 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 661
           | || ||||||||| |||||||||| || ||||| ||||| ||||| || || || ||||
Db    1034 AACTTTACAATGCTCTTCTGGCTGGATCAGTACCAGTTGTACTGGGCCCCTCCAGAGAAA 1093

Qy     662 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 721
           |||| |||||||| |||||||||| || || || || |||||||||| |   || || |
Db    1094 ACTACGAGAATTACATTCCAGCAGACTCTTTTATCCACGTGGAAGATTTTCTATCCCCAA 1153

Qy     722 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 781
           | ||||| ||| | ||||| |   | |||||||| || ||| | ||||||||||||||
Db    1154 GAGAGCTTGCAGAATATCTTCTGATGCTTGACAAAAATAACAAGATGTACCTTAGTTACT 1213

Qy     782 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 841
           | |||||||||||||||| | ||  |||||||   ||| |||||||| |||||||||
Db    1214 TCAACTGGAGGAAGGATTTCTCAGTGCATCTTCCTAAATTCTGGGAATCTCATGCATGTC 1273

Qy     842 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAAT 901
           | ||||| |||||||||||||| || || ||||| |||||  |||| |||||||| ||||
Db    1274 TTGCTTGTGATCATGTGAAAAGACACCAGGAATACAAGTCAATTGGAAATTTAGAAAAAT 1333

Qy     902 GGTTTTGGAATTAA 915
           |||||||||| |||
Db    1334 GGTTTTGGAACTAA 1347


RESULT 9
AB035905
LOCUS       AB035905               1080 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFUT9, complete cds.
ACCESSION   AB035905
VERSION     AB035905.1  GI:18146863
KEYWORDS    xFUT9.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
COMMENT     Sequence updated (29-Feb-2000).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .1080
                     /gene="xFuc-TIX"
     CDS             1. .1080
                     /gene="xFuc-TIX"
                     /codon_start=1
                     /product="xFUT9"
                     /protein_id="BAB82488.1"
                     /db_xref="GI:18146864"
                     /translation="MTSASKGILRAFLAVSGVLICFMICLLIYVKPTNNWISSPIESA
                     HSVLKMKNFFSSKQNYYNETVVLIWVWPFGQTFELKSCQTLFNIYGCHLTTDRALYNK
                     SHAVLLHHRDISWDLINLPQQLRPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRR
                     DSDIQVPYGFMVVSTKPFDFEVPSKDKFVCWVVSNWNPEHARVKYYNELIKYIEIITY
                     GQAFGEYLSDKSLLPTIASCKFYLSFENSIHKDYITEKLYNALLAGSVPIVLGPPREN
                     YENYIPADSFIHVEDFLSPRELADHLLMLNKDTEQYLSYYNWRKHFTVNMPHFWESHA
                     CLACDHVKRHQEYKSVGNLEKWFWN"
BASE COUNT      332 a    231 c    202 g    315 t
ORIGIN

  Query Match           63.5%;  Score 581.2;  DB 5;  Length 1080;
  Best Local Similarity   77.2%;  Pred. No. 1e-142;
  Matches  706;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy       2 CCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGA 61
           | |||    ||||||  ||||||||     ||||  | ||||||||||| ||||||||||
Db     167 CTAAACAGAATTATTACAATGAAACCGTCGTTCTAATTTGGGTGTGGCCTTTTGGGCAGA 226

Qy      62 CCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGG 121
           | ||||| || | |||||||||  || |||| || ||  | || |||||||| || || |
Db     227 CTTTTGAACTGAAATCCTGCCAGACATTGTTTAATATTTATGGGTGCCATCTTACTACTG 286

Qy     122 ACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGG 181
           ||||  |||| |||||||| || ||||| || ||  | || ||| ||||||| || ||||
Db     287 ACCGCGCACTCTACAACAAGTCTCATGCTGTCCTTCTTCACCACAGAGACATTAGCTGGG 346

Qy     182 ATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATT 241
           |||| |  ||  | || || ||    |||||||| ||||||||||||||||||||||||
Db     347 ATCTCATTAACCTGCCCCAACAGCTGAGGCCACCTTTCCAGAAATGGATTTGGATGAATC 406

Qy     242 TGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTC 301
           |||||||||||| || || || ||||| |||||||||||||||| ||||||||||||
Db     407 TGGAATCACCAACACATACGCCACAAAAAAGTGGCATTGAGCACTTATTTAACCTGACTC 466

Qy     302 TGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAA 361
           |||||||| | ||||| |||||||||||||||||||||||||||| ||   ||||||||||
Db     467 TGACTTACAGGCGTGACTCAGATATCCAAGTGCCTTATGGCTTCATGGTCGTAAGCACAA 526

Qy     362 ATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACT 421
           | || || |  ||||| |||||||||||||| || || || || |||||||| || ||||
Db     527 AACCATTTGACTTTGAGGTGCCAAGCAAAGACAAGTTTGTCTGTTGGGTTGTTAGCAACT 586
```

```
Qy     422 GGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCC 481
           ||||||| || |||||||||||||| |||||||||| || | ||||  |||||||||
Db     587 GGAACCCAGAACATGCCAGAGTCAAATATTACAATGAACTTATCAAATACATTGAAATTA 646

Qy     482 ATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATAT 541
           || || |||||||| ||||| |||||  | |  |||||||   || |||| || ||
Db     647 TAACTTATGGGCAAGCTTTTGGTGAATACCTGAGCGATAAAAGCCTGCTTCCAACTATTG 706

Qy     542 CTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAA 601
           |  |||| |||||||||||| || |||||||||||| || ||||| ||||| ||||| ||||
Db     707 CATCTTGCAAATTTTATCTGTCTTTTGAAAATTCCATTCACAAAGATTATATCACAGAAA 766

Qy     602 AGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAA 661
           | || || ||||| ||  ||||||| || || ||  ||||  |||||||  |||| ||||
Db     767 AACTTTATAATGCATTATTGGCTGGGTCAGTGCCCATTGTATTGGGACCTCCTAGAGAAA 826

Qy     662 ACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCA 721
           ||||||||| || ||||| |||||||| || || || |||||||||  |  |||||| |
Db     827 ACTATGAGAACTACATTCCTGCAGATTCCTTTATCCACGTGGAAGATTTTCTCTCTCCAA 886

Qy     722 GTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACT 781
           | ||  ||||  |  |||         |  ||||| | |   |    ||||| || ||||
Db     887 GAGAATTAGCTGACCATCTTTTAATGCTGAACAAAGATACAGAACAGTACCTGAGCTACT 946

Qy     782 TTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTT 841
           ||||||||||  |||| || || ||| | || |  || |||||||| |||||||| |
Db     947 ACAACTGGAGGAAACATTTTACAGTCAATATGCCTCACTTCTGGGAATCCCATGCATGCT 1006

Qy     842 TGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAAT 901
           | || || || |||||||||||||||||||||| || || || || ||| |||||||| |
Db    1007 TAGCGTGTGACCATGTGAAAAGGCATCAAGAATACAAATCAGTCGGGAATCTAGAGAAGT 1066

Qy     902 GGTTTTGGAATTAA 915
           |||| ||||| |||
Db    1067 GGTTCTGGAACTAA 1080
```

```
RESULT 10
AC100674
LOCUS       AC100674               71396 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-167K24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100674
VERSION     AC100674.1  GI:17048040
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-167K24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15932
                Center clone name: 167_K_24
            --------------.
            * NOTE: This record contains 71 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1702: contig of 814 bp in length
            *     1703 1802: gap of      100 bp
            *     1803     2642: contig of 840 bp in length
            *     2643 2742: gap of      100 bp
            *     2743     3637: contig of 895 bp in length
            *     3638 3737: gap of      100 bp
            *     3738     4635: contig of 898 bp in length
            *     4636 4735: gap of      100 bp
            *     4736     5657: contig of 922 bp in length
            *     5658 5757: gap of      100 bp
            *     5758     6692: contig of 935 bp in length
            *     6693 6792: gap of      100 bp
            *     6793     7703: contig of 911 bp in length
            *     7704 7803: gap of      100 bp
            *     7804     8742: contig of 939 bp in length
            *     8743 8842: gap of      100 bp
            *     8843     9787: contig of 945 bp in length
            *     9788 9887: gap of      100 bp
            *     9888    10819: contig of 932 bp in length
            *    10820 10919: gap of      100 bp
            *    10920    11808: contig of 889 bp in length
            *    11809 11908: gap of      100 bp
            *    11909    12830: contig of 922 bp in length
            *    12831 12930: gap of      100 bp
            *    12931    13847: contig of 917 bp in length
            *    13848 13947: gap of      100 bp
            *    13948    14886: contig of 939 bp in length
            *    14887 14986: gap of      100 bp
            *    14987    15911: contig of 925 bp in length
            *    15912 16011: gap of      100 bp
            *    16012    16903: contig of 892 bp in length
            *    16904 17003: gap of      100 bp
            *    17004    17906: contig of 903 bp in length
            *    17907 18006: gap of      100 bp
            *    18007    18931: contig of 925 bp in length
            *    18932 19031: gap of      100 bp
            *    19032    19945: contig of 914 bp in length
            *    19946 20045: gap of      100 bp
            *    20046    20951: contig of 906 bp in length
            *    20952 21051: gap of      100 bp
            *    21052    21949: contig of 898 bp in length
            *    21950 22049: gap of      100 bp
            *    22050    22950: contig of 901 bp in length
```

```
*    22951 23050: gap of      100 bp
*    23051    23945: contig of 895 bp in length
*    23946 24045: gap of      100 bp
*    24046    24955: contig of 910 bp in length
*    24956 25055: gap of      100 bp
*    25056    25980: contig of 925 bp in length
*    25981 26080: gap of      100 bp
*    26081    27052: contig of 972 bp in length
*    27053 27152: gap of      100 bp
*    27153    28067: contig of 915 bp in length
*    28068 28167: gap of      100 bp
*    28168    29076: contig of 909 bp in length
*    29077 29176: gap of      100 bp
*    29177    30093: contig of 917 bp in length
*    30094 30193: gap of      100 bp
*    30194    31099: contig of 906 bp in length
*    31100 31199: gap of      100 bp
*    31200    32097: contig of 898 bp in length
*    32098 32197: gap of      100 bp
*    32198    33113: contig of 916 bp in length
*    33114 33213: gap of      100 bp
*    33214    34125: contig of 912 bp in length
*    34126 34225: gap of      100 bp
*    34226    35133: contig of 908 bp in length
*    35134 35233: gap of      100 bp
*    35234    36150: contig of 917 bp in length
*    36151 36250: gap of      100 bp
*    36251    37171: contig of 921 bp in length
*    37172 37271: gap of      100 bp
*    37272    38159: contig of 888 bp in length
*    38160 38259: gap of      100 bp
*    38260    39156: contig of 897 bp in length
*    39157 39256: gap of      100 bp
*    39257    40188: contig of 932 bp in length
*    40189 40288: gap of      100 bp
*    40289    41210: contig of 922 bp in length
*    41211 41310: gap of      100 bp
*    41311    42222: contig of 912 bp in length
*    42223 42322: gap of      100 bp
*    42323    43218: contig of 896 bp in length
*    43219 43318: gap of      100 bp
*    43319    44233: contig of 915 bp in length
*    44234 44333: gap of      100 bp
*    44334    45246: contig of 913 bp in length
*    45247 45346: gap of      100 bp
*    45347    46261: contig of 915 bp in length
*    46262 46361: gap of      100 bp
*    46362    47279: contig of 918 bp in length
*    47280 47379: gap of      100 bp
*    47380    48276: contig of 897 bp in length
*    48277 48376: gap of      100 bp
*    48377    49282: contig of 906 bp in length
*    49283 49382: gap of      100 bp
*    49383    50305: contig of 923 bp in length
*    50306 50405: gap of      100 bp
*    50406    51314: contig of 909 bp in length
*    51315 51414: gap of      100 bp
*    51415    52335: contig of 921 bp in length
*    52336 52435: gap of      100 bp
*    52436    53163: contig of 728 bp in length
*    53164 53263: gap of      100 bp
*    53264    54210: contig of 947 bp in length
*    54211 54310: gap of      100 bp
*    54311    55199: contig of 889 bp in length
*    55200 55299: gap of      100 bp
*    55300    56212: contig of 913 bp in length
*    56213 56312: gap of      100 bp
*    56313    57287: contig of 975 bp in length
*    57288 57387: gap of      100 bp
*    57388    58245: contig of 858 bp in length
*    58246 58345: gap of      100 bp
*    58346    59236: contig of 891 bp in length
*    59237 59336: gap of      100 bp
*    59337    60272: contig of 936 bp in length
*    60273 60372: gap of      100 bp
*    60373    61289: contig of 917 bp in length
*    61290 61389: gap of      100 bp
*    61390    62299: contig of 910 bp in length
*    62300 62399: gap of      100 bp
*    62400    63322: contig of 923 bp in length
*    63323 63422: gap of      100 bp
*    63423    64336: contig of 914 bp in length
*    64337 64436: gap of      100 bp
*    64437    65354: contig of 918 bp in length
*    65355 65454: gap of      100 bp
*    65455    66352: contig of 898 bp in length
*    66353 66452: gap of      100 bp
*    66453    67362: contig of 910 bp in length
*    67363 67462: gap of      100 bp
*    67463    68370: contig of 908 bp in length
*    68371 68470: gap of      100 bp
*    68471    69379: contig of 909 bp in length
*    69380 69479: gap of      100 bp

 Query Match             63.4%;  Score 580.2;  DB 2;  Length 71396;
 Best Local Similarity   86.0%;  Pred. No. 1.8e-142;
 Matches  692;  Conservative    0;  Mismatches  100;  Indels   13;  Gaps    4;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
           || |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db   65550 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 65609

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           |||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db   65610 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 65669

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db   65670 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 65729

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||||| || |||||||||||| ||||||||||||||||||||
Db   65730 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 65789

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db   65790 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 65849

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   65850 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 65909

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db   65910 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 65969

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           |||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db   65970 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 66029

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db   66030 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 66089

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| ||||||||||||||||||  |  ||| |       ||  |  | | |   | |
Db   66090 TCTACTTGTAAATTTTATCTTTCTCTGAAAACTCATTCACAAGATTACTCCCAAAAG--- 66146

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
                 | ||     ||| ||||||| || |||||||||| ||||| ||||||||||||
Db   66147 ------TCCACATGCTTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 66200

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTC--ATGTGGAAGATTATAACTCT- 717
           |||||||||||||||||||||||| ||||||||||||    |||||||||| | |||||
Db   66201 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCCATCGTGGAAGATTTTCACTCTC 66260
```

```
Qy     718 CCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTT-ATACCTTAG 776
           ||||||||| ||||||| ||||||||||||||||| |||||||||| ||||||||
Db   66261 CCCAGTGAGTTAGCAAATATCCTGAAGGAAGTCGACAANAACAATAAGTTGGTACCTTAG 66320

Qy     777 TTACTTTAACTGGAGGAAGGATTTC 801
           ||||| ||||| |  ||| |  |||
Db   66321 TTACTCTAACTCGNAGAAAGGATTC 66345


RESULT 11
AB023627
LOCUS       AB023627                1567 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT1 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023627
VERSION     AB023627.1  GI:4587295
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1567)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            177. .1352
                     /gene="zFT1"
     CDS             177. .1352
                     /gene="zFT1"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76706.1"
                     /db_xref="GI:4587296"
                     /translation="MDKITLTPPSKAAQKVIIATFMLISFVCIFFVYYNPNTTFFKFP
                     AVDVHTNCSCTETCLDVLKKQNYKCTIKNASDNPSTTPAPNPAPATAKKDDQETILLI
                     WVYPFGARFDLGVCGSQFNIHGCHLTDDRSMIQKAHGVMFHHRDLSADLPQPPRPAFQ
                     KWIWWNMESPSNSHPNSLINDGFNLTSSFRKDSDIPVPYGRLIDATDEQKNFTIPKKD
                     KLVCWIVSNFQEHHKRSQYYNELVKHIKVEAYGGHFNNRVNDDDYGNVVSSCKFYLSF
                     ENSIHRDYFTEKLFNPLALGTVPVVLGPSRDNYEEFIPRDAFIHVDDFPTPKELADHL
                     KSLDQNEEQYRQYFNWRKHFISMSSSFGLEHACRTCDYLKRNKHYIAIKDLKGWFWG"
BASE COUNT      499 a    334 c    293 g    441 t
ORIGIN

  Query Match             30.4%;  Score 277.8;  DB 5;  Length 1567;
  Best Local Similarity   58.5%;  Pred. No. 1.2e-62;
  Matches  527;  Conservative    0;  Mismatches  362;  Indels   12;  Gaps    2;

Qy      18 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 77
           | | ||||| |   |  || | |||||||  || || ||       ||||||  |   |
Db     461 TCAAGAAACCATCCTGTTGATCTGGGTGTATCCTTTCGGAGCACGTTTTGACTTGGGAGT 520

Qy      78 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 137
             ||     |  |||||||||||| || || ||| | |||  ||| |    ||  | |
Db     521 GTGTGGTTCGCAGTTCAACATCCACGGCTGTCATTTAACAGACGACAGAAGTATGATCCA 580

Qy     138 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 197
           ||| | |||| |||| || | || ||| ||||  | ||  ||          |||| ||
Db     581 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATCTGAGCGCGG---------ATTTGCC 631

Qy     198 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 257
           || |   || | ||| | ||||| || ||||| |||  |||| |||| || || | | |
Db     632 GCAACCCCCTCGACCAGCTTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGAGTAA 691

Qy     258 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 317
             | |  | |||  ||   ||  |  |  |||||| |||||    | || ||||  ||
Db     692 TTCCCATCCAAACAGCTTGATAAATGATGGGTTTAATCTGACGTCAAGTTTCCGCAAGGA 751

Qy     318 TTCAGATATCCAAGTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTT 374
           |||||||||  |||||||||    | |  || |   |  | |    | ||       ||
Db     752 TTCAGATATTCCAGTGCCTTACGGTCGACTCATCGATGCCACCGACGAACAGAAAAACTT 811

Qy     375 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 434
             | ||  | ||  || ||| | || |||||| | || |||||||    |     | ||
Db     812 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGAACACCA 871

Qy     435 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 494
               |    |||| || |||||| |    |||  |||| || |  |  |||| ||
Db     872 CAAACGTTCACAGTACTATAATGAGTTGGTGAAACACATTAAAGTGGAGGCCTATGGAGG 931

Qy     495 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 554
              ||      |  || |||||| |  |||      |    ||||   |||||||||
Db     932 GCACTTCAATAACCGGGTTAATGATGACGATTATGGAAATGTGGTATCCAGTTGTAAATT 991

Qy     555 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 614
             || |||||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db     992 CTACCTTTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1051

Qy     615 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 674
           |||    |  || | || |||||||| || || || || || || |||||||||  | |
Db    1052 TTTAGCGCTCGGTACGGTTCCTGTTGTACTTGGCCCTTCCAGAGATAACTATGAGGAGTT 1111

Qy     675 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 734
           |||  |||  |||  | ||||| |||||||| ||||       ||||||  || || ||  |
Db    1112 TATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1171

Qy     735 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 794
             | ||  ||    | ||| ||||| |  |    |||       || || |||||||| ||
Db    1172 TCACCTTAAATCCTTGGACCAAAACGAGGAACAGTACAGACAGTATTTCAACTGGAGAAA 1231

Qy     795 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 854
             |||| | |  ||    | ||  |||| ||      ||||| |||  | |||| ||| |
Db    1232 ACATTTTATTTCAATGAGTTCATCATTTGGGCTTGAACATGCCTGTAGGACTTGTGATTA 1291

Qy     855 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 914
           | | |||||| || |  | ||||   || ||    || |  |    ||||| |||  |||
Db    1292 TCTTAAAAGGAATAAGCACTATATAGCTATTAAAGATCTGAAAGGCTGGTTCTGGGGTTA 1351

Qy     915 A 915
           |
Db    1352 A 1352


RESULT 12
AB023628
LOCUS       AB023628                1832 bp    DNA     linear   VRT 02-APR-1999
DEFINITION  Danio rerio zFT2 gene for alpha(1,3)fucosyltransferase, complete
            cds.
ACCESSION   AB023628
VERSION     AB023628.1  GI:4587297
KEYWORDS    alpha(1,3)fucosyltransferase.
SOURCE      Danio rerio DNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Molecular cloning and characterization of two zebrafish
            alpha(1,3)fucosyltransferase genes developmentally regulated in
            embryogenesis
  JOURNAL   J. Biochem. (1999) In press
REFERENCE   2  (bases 1 to 1832)
```

```
  AUTHORS   Kageyama,N., Natsuka,S. and Hase,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Shunji Natsuka, Osaka University, Graduate
            School of Science; Machikaneyama 1-1, Toyonaka, Osaka 560-0034,
            Japan (E-mail:natsuka@chem.sci.osaka-u.ac.jp, Tel:81-6-6850-5381,
            Fax:81-6-6850-5383)
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            332. .1510
                     /gene="zFT2"
     CDS             332. .1510
                     /gene="zFT2"
                     /codon_start=1
                     /product="alpha(1,3)fucosyltransferase"
                     /protein_id="BAA76707.1"
                     /db_xref="GI:4587298"
                     /translation="MDKITVMDKIKLTAPSKSTQKLILVTFMLISLVCIFYVYYNPNT
                     TFFKYPTMGIHANSSCSETCLHVLKMQNYECIMKNASYNALPTPAPNPAKGDQDIIVL
                     IWMAPFRVPFDLKDCGLEYNIHGCQLTYDRSLIQKAHGVMFHHRDMSTDFPQPPRPDF
                     QKWIWWNMESPTHCNQNDFLNDRFNVTSSYNRDSDIPVPYGRLVDATEEQKKFTIPKK
                     DKLVCWIVSNFQGHHKRTHFFYEFAKHINVSTYGGHFNNRINGDDYGNVVSSCKFYLS
                     FENSIHRDYFTEKLFNPLALGTVPVVIGPPRYNYEEFIPRDAFIHVDDFPTPKELADH
                     LISLDQNEEQYKQFFNWRKLYVSKSTSFGLEHSCRICDYLKRNKHYIAVTDLKGWFWG
                     "
BASE COUNT      585 a    394 c    317 g    536 t
ORIGIN

  Query Match           25.6%;  Score 234.6;  DB 5;  Length 1832;
  Best Local Similarity   55.5%;  Pred. No. 2.9e-51;
  Matches  500;  Conservative    0;  Mismatches  389;  Indels   12;  Gaps    2;

Qy      18 TAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATC 77
           | | || |  |   |  || | ||| ||   || ||  |      | ||||||| | |
Db     619 TCAAGACATCATCGTGTTGATCTGGATGGCTCCTTTCAGAGTACCTTTTGACCTGAAAGA 678

Qy      78 CTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAA 137
            ||         ||  |||||||| || || ||  | |||   ||| |    ||| | |
Db     679 TTGTGGTTTGGAGTATAACATCCACGGCTGTCAATTAACATACGACAGAAGTCTGATCCA 738

Qy     138 CAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACC 197
           ||| | |||| |||| || | || ||| |||| || ||           ||| |||| ||
Db     739 GAAAGCGCATGGAGTTATGTTTCACCACAGAGATATGAGC---------ACAGATTTTCC 789

Qy     198 TCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCA 257
           || |   |  | |||   ||||| || ||||| |||  |||| |||| || || || ||
Db     790 GCAACCCCCGCGACCAGATTTCCAAAAGTGGATATGGTGGAATATGGAGTCTCCGACCCA 849

Qy     258 CACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGA 317
             |||||      | |  |  |   ||||||  ||||    | ||||  | | ||
Db     850 TTGCAATCAAAATGACTTCCTAAATGATCGGTTTAATGTGACGTCAAGTTACAACAGGGA 909

Qy     318 TTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA---AATCCCTTCGTGTT 374
           |||||||||| | |||||||||| ||      |    |    ||||    | |       |||
Db     910 TTCAGATATTCCAGTGCCTTACGGTCGACTCGTCGATGCCACAGAGGAACAGAAAAAGTT 969

Qy     375 TGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCA 434
           | || |  || || ||| | || ||||||  | || |||||||    |     | ||
Db     970 CACCATCCCCAAAAAGGATAAACTAGTTTGCTGGATAGTAAGTAACTTTCAAGGGCACCA 1029

Qy     435 TGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCA 494
           |      | |  | | | ||| |    |||  |||| |  |    ||||| ||
Db    1030 CAAGCGCACACATTTCTTCTACGAGTTTGCGAAACACATTAATGTGAGCACCTATGGAGG 1089

Qy     495 AGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATT 554
           ||      |      | |||| | |  |||        |  | ||||   |||||||||
Db    1090 GCACTTCAACAATCGAATTAATGGTGACGATTATGGAAATGTCGTATCCAGTTGTAAATT 1149

Qy     555 TTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGC 614
           || || |||||||| || || || ||||  || ||  |||| || ||||| | |||| |
Db    1150 CTACCTGTCCTTTGAGAACTCCATTCACAGAGACTATTTCACTGAGAAGCTGTTCAATCC 1209

Qy     615 TTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTA 674
           |||       ||  | || ||||| ||  | || ||  | ||  | |||||||| | |
Db    1210 TTTAGCACTCGGTACAGTTCCTGTCGTAATTGGCCCGCCCAGATATAACTATGAAGAGTT 1269

Qy     675 TATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAA 734
           || |||  |||  | ||||| |||||||| ||||      ||||||  || || ||  |
Db    1270 CATACCAAGAGACGCCTTCATCCATGTGGATGATTTCCCAACTCCCAAGGAACTGGCGGA 1329

Qy     735 GTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAA 794
           | || |      | ||| ||||| |  |    |||       |  || ||||||||||
Db    1330 TCACCTTATATCCTTGGACCAAAACGAGGAACAGTACAAACAGTTTTTCAACTGGAGGAA 1389

Qy     795 GGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCA 854
           ||    |  |||   | ||   ||| ||      |||| | |||  |  ||| ||| |
Db    1390 ACTTTATGTTTCAAAGAGTACATCGTTTGGGCTTGAACATTCCTGTAGGATTTGTGATTA 1449

Qy     855 TGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATTA 914
           | | |||||| || |  | ||||   | |||    || |  |    |||| ||| |||
Db    1450 TCTTAAAAGGAATAAGCACTATATAGCCGTTACGGATCTGAAAGGGTGGTTCTGGGGTTA 1509

Qy     915 A 915
           |
Db    1510 A 1510


RESULT 13
AB035907
LOCUS       AB035907               1092 bp    DNA     linear   VRT 05-JAN-2002
DEFINITION  Xenopus laevis gene for xFTI, complete cds.
ACCESSION   AB035907
VERSION     AB035907.1  GI:18146867
KEYWORDS    xFTI.
SOURCE      Xenopus laevis bone marrow DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Kitano,T., Narimatsu,H. and Saitou,N.
  TITLE     The evolutionary history of glycosyltransferase genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Kaneko,M., Saitou,N. and Kitano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Mika Kaneko, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; Yata 1111, Mishima,
            Shizuoka 411-8540, Japan (E-mail:mkaneko@med.id.yamagata-u.ac.jp,
            Tel:81-559-81-6790, Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="bone marrow"
     gene            1. .1092
                     /gene="xFTI"
     CDS             1. .1092
                     /gene="xFTI"
                     /codon_start=1
                     /product="xFTI"
                     /protein_id="BAB82490.1"
                     /db_xref="GI:18146868"
                     /translation="MRGDSLLSVERLRCWRLCRRSRLRSLRSRLLAASLTCTVLVVYV
                     CSQDLIAGPINFPLLWGTTTPPKQVTVLIWYEPFGKRRRLGDCHMLFNVTGCHITTNR
                     SLYQEADAIMFHHRDISDFSDLPFRRRQSSQKWIWMNFESPSHSPWLASLGGIFNWTM
                     SYRVDSDIFMPYGFLFSKKRAKIVLPRKRKLVAWVISNWNEDHERVQYYNELRNYVEI
                     DVYGRYGLDLKEDNIVKTVSEYKFYLAFENSLHTDYITEKLWRNAFKSNAIPIVMGPS
                     RYNYEMFIPRSSFIHVDDFSSPRKLAMYLKHLDKNIHLYRRYFTWKKRYDVHVTSFWD
                     EHYCTACQSVKAAGNQHRTASDLAGWFES"
BASE COUNT      256 a    281 c    294 g    261 t
ORIGIN

  Query Match           24.3%;  Score 222.6;  DB 5;  Length 1092;
  Best Local Similarity   57.3%;  Pred. No. 4.2e-48;
```

```
  Matches  484;  Conservative    0;  Mismatches  349;  Indels   12;  Gaps    4;

Qy      19 AATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCC 78
           ||  ||   ||| |  |  | |||     |||||||||  |          |||
Db     199 AAGCAAGTCACTGTCTTAATATGGTATGAGCCATTTGGCAAACGGCGAAGGCTTGGGGAT 258

Qy      79 TGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAAC 138
           |||||     | |||||  ||   || |||||  |||||||  || |    || ||| |
Db     259 TGCCACATGCTCTTCAATGTCACGGGGTGCCACATCACAACCAACAGGAGTCTTTACCAG 318

Qy     139 AAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCT 198
           |  |  | ||  || || |||| ||||| |||||   ||  ||| | |   |  | |||
Db     319 GAGGCAGACGCCATTATGTTCCACCACCGGGACAT---TTCTGATTTCAGTGACCTTCCT 375

Qy     199 CAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 258
           | |    | |   || ||| || ||||| |||||||| || |||||  ||  | ||
Db     376 TTCCGAAGGCGCCAGAGCTCCCAAAAGTGGATCTGGATGAACTTCGAATCTCCGTCCCAT 435

Qy     259 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 318
           | |||     |    |  | | |  | | || |||  |||  || | ||| |    ||
Db     436 TCGCCCTGGCTGGCGAGTCTCGGGGGCATCTTCAACTGGACCATGTCCTACAGGGTGGAC 495

Qy     319 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 378
           || ||||||    |||| || ||||||||      |   ||  || |    |    | |
Db     496 TCGGATATCTTCATGCCCTACGGCTTCTT---ATTCTCCAAGAAACGTGCCAAAATAGTC 552

Qy     379 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 438
           ||||  ||||    ||  |||||  ||||||  | || ||||||||||   || || |
Db     553 CTGCCGCGCAAGAGAAAGCTGGTGGCCTGGGTGATCAGCAACTGGAACGAGGACCACGAA 612

Qy     439 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 498
           || ||| |||||||||| ||||| || ||   | | ||||||  |    |||||  |
Db     613 AGGGTCCAGTATTACAACGAGCTGAGGAACTACGTCGAAATCGACGTTTACGGCC---GG 669

Qy     499 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 558
           | |||     || | || ||  | ||  |  |    ||| | || |  | |||||||||
Db     670 TATGGGCTGGATTTAAAAGAGGACAACATCGTCAAGACCGTCTCGGAGTATAAATTTTAC 729

Qy     559 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 615
           |  | ||||| |||||  | |||| |||||||||||||||||||||  |      || ||
Db     730 TTGGCTTTTGAGAATTCCTTGCACACGGATTACATCACGGAAAAGCTGTGGAGGAACGCG 789

Qy     616 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 675
           |||    |  | | || |  || |||| ||      |  | ||||| |||| |
Db     790 TTTAAATCCAACGCCATCCCCATCGTCATGGGGCCCAGCCGCTATAACTACGAGATGTTC 849

Qy     676 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 735
           || ||       || |||||||| ||||| ||||    |   |||||  |||| || | |
Db     850 ATACCGCGCAGCTCTTTCATTCACGTGGACGATTTCTCCAGCCCCAGGAAGCTGGCCATG 909

Qy     736 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
           || |||||| |  | || ||||||| | | || ||||  || || || |  |||| ||||
Db     910 TACCTGAAGCACCTGGATAAAAACATTCACTTGTACCGCAGGTATTTCACGTGGAAGAAG 969

Qy     796 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 855
             || |   || |  |  |    || |||||    |||   ||    || ||| |
Db     970 CGTTACGACGTGCACGTGACGTCTTTCTGGGACGAGCATTACTGCACCGCGTGCCAGAGC 1029

Qy     856 GTGAA 860
           || ||
Db    1030 GTCAA 1034


RESULT 14
AF288369
LOCUS       AF288369               1264 bp    DNA     linear   VRT 02-DEC-2001
DEFINITION  Gallus gallus alpha (1,3) fucosyltransferase gene, complete cds.
ACCESSION   AF288369
VERSION     AF288369.1  GI:17224489
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Expression of fucosyltransferases in chick embryos
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1264)
  AUTHORS   Foley,A.C., Schmitz,B., Stern,C.D. and Streit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Genetics and Development, Columbia
            University, 701 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     CDS             <1. .1062
                     /note="fut"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /protein_id="AAL36994.1"
                     /db_xref="GI:17224490"
                     /translation="SWKKLLTFLLFSFIFSFCLFASLRQFRTPEPEDYNSYLSTERLP
                     AGSTSMLPKSERMLTILLWTWPFGHRFNFSNCSELFDTPNCHFTTNRSWSQKADAVIM
                     HHRDVYWDRERLSRIPRLPSQRWIWFNMESPSHSPNLGAMDHLFNLTMSYRRDSDIFT
                     PYGELQLLSEPQPFSIPPKTKMVAWVVSNWRPSSHRVKYYEELKKHIPVDVYGRQHLP
                     LSMDMLHQTVSQYNFYLAFENSEHKDYITEKLWRNALSSGTIPVVLGPPRENYEHFLP
                     PDSFIHVDDFASAGELAQYLQELSRDTEKYQRYFQWRRWLKPVLGIGWAMQVCRACHF
                     LQRTEASYHVVPNLSKWFV"
BASE COUNT      286 a    355 c    319 g    304 t
ORIGIN

  Query Match             24.1%;  Score 220.2;  DB 5;  Length 1264;
  Best Local Similarity   58.0%;  Pred. No. 1.8e-47;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    6;  Gaps    3;

Qy      28 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 87
           || || ||| |||||   |||||  || || ||   |||| |  | |    ||||  ||
Db     175 ACCATCCTGCTGTGGACATGGCCCTTCGGTCACCGCTTTAATTTCAGCAACTGCTCAGAG 234

Qy      88 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             | ||  ||| ||      |||||  |||| ||  ||||       || | | ||  |  ||
Db     235 CTCTTTGACACCCCGAACTGCCACTTCACCACCAACCGCAGCTGGTCCCAAAAGGCAGAT 294

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCT 207
           || ||  | || ||||||  | ||  |   ||||||       |   |  | | |    |
Db     295 GCTGTGATTATGCATCACAGGGATGTGTATTGGGACAGAGAGAGGCTGTCCCGGATCCCC 354

Qy     208 AGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAA 267
           ||||  || | ||||   ||||| ||| | ||  |||| || |||| |||  ||||  |
Db     355 AGGCTCCCATCCCAGCGCTGGATCTGGTTCAACATGGAGTCCCCAAGTCATTCTCCAAAC 414

Qy     268 AAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATC 327
             ||| ||| || ||| | || |||||||||| || | |||||  | || || || |||
Db     415 TTAGGTGCCATGGACCACCTCTTCAACCTGACTATGTCGTACCGGAGAGACTCGGACATC 474

Qy     328 CAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 387
           .    || |||||    ||  | |   ||   | |||      ||     | |||  |
Db     475 TTCACCCCCTATGGGGAGCTGCAGCTGCTCAGCGAGCCCCAGCCCTTCAGCATCCCACCC 534

Qy     388 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAG 447
           ||    ||  ||||| |||||| || |||| |||| ||           | | || |||
Db     535 AAGACCAAGATGGTGGCCTGGGTGGTCAGTAATTGGAGACCAAGCTCCCACCGGGTGAAG 594

Qy     448 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 507
           || ||| | ||||| |  |||  |||      |  ||   |||||||  |||      |
Db     595 TACTACGAGGAGCTGAAGAAACACATCCCTGTGGATGTGTACGGGC-GGCAACACTTGCC 653

Qy     508 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 567
             |||| |||   | |   |    |  ||| ||||     |  || || ||  |  |||||
Db     654 CCTGTCCATG--GACATGCTCCACCAAACCGTATCCCAGTACAACTTCTACTTGGCCTTT 711

Qy     568 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATAC---AATGCTTTTCTGGCT 624
           || || |||   ||||| || |||||||| || ||||| |     |||||  |     ||
```

```
Db    712 GAGAACTCAGAGCACAAAGACTACATCACTGAGAAGCTCTGGAGGAATGCCCTGTCCTCT 771

Qy    625 GGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCA 684
          ||| |  | || |||||||||||| ||| || | |||||||||||| | |   | ||  |
Db    772 GGCACCATCCCAGTTGTTCTGGGGCCACCTCGAGAAAACTATGAGCACTTCTTGCCCCCT 831

Qy    685 GATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAG 744
          || || ||||| || || || ||-| |  |  | |  |||| || ||  |||| ||| ||
Db    832 GACTCCTTCATCCACGTCGATGACTTTGCCAGTGCTGGTGAACTGGCGCAGTACCTGCAG 891

Qy    745 GAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          ||| |   ||   |||   ||   ||||   | |||||  | ||| | | |
Db    892 GAACTGAGCAGGGACACCGAGAAGTACCAGCGCTACTTCCAGTGGCGCAGG 942


RESULT 15
GGU73678
LOCUS       GGU73678               2657 bp    DNA     linear   VRT 20-DEC-1996
DEFINITION  Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete
            cds.
ACCESSION   U73678
VERSION     U73678.1  GI:1657998
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N.,
            Schultz,T.L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
  TITLE     Molecular cloning and characterization of CFT1, a developmentally
            regulated avian alpha(1,3)-fucosyltransferase gene
  JOURNAL   J. Biol. Chem. 271 (51), 32960-32967 (1996)
  MEDLINE   97115837
REFERENCE   2  (bases 1 to 2657)
  AUTHORS   Lee,K.P., Carlson,L.M. and Larsen,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical
            Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES             Location/Qualifiers
     source          1. .2657
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            170. .1240
                     /gene="CFT1"
     CDS             170. .1240
                     /gene="CFT1"
                     /note="similar to mammalian fucosyltransferase FucTIV"
                     /codon_start=1
                     /product="alpha-1,3-fucosyltransferase"
                     /protein_id="AAC60060.1"
                     /db_xref="GI:1657999"
                     /translation="MELGPRWSPAARPGCPRRWRRRWALLGALLGAALALYVCVRELR
                     RRGSAAGRPEGEVTVLLWWEPFGRPWRPADCRRRYNITGCLLSADRGRYGEARAVLFH
                     HRDLALHGRQGLPRGPPPRPPRQRWVWMNFESPSHSPGLRGLAGLFNWTMSYRRDSDV
                     FVPYGYLYEPPSPRPFVLPRKSRLVAWVISNWNEEHARVRYYRQLKEHLPIDVYGARG
                     MALLEGSVVKTVSAYKFYLAFENSQHTDYITEKLWKNAFAASAVPVVLGPRRANYERF
                     IPADSFIHVDDFPSPRLLATYLKFLDKNKPSYRRYFAWRNKYEVHVTSFWDEHYCKVC
                     EAVRTAGNQLKTVQNLAGWFES"
BASE COUNT      581 a    712 c    818 g    546 t
ORIGIN

  Query Match           19.3%;  Score 176.2;  DB 5;  Length 2657;
  Best Local Similarity  53.1%;  Pred. No. 7.3e-36;
  Matches  451;  Conservative    0;  Mismatches  383;  Indels   15;  Gaps    3;

Qy     28 ACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCA 87
          ||  | ||| |||||  |  ||| || || |   |||    ||   |  |||||
Db    341 ACGGTGCTGCTGTGGTGGGAGCCCTTCGGCCGCCCCTGGCGCCCGGCCGACTGCCGGAGG 400

Qy     88 ATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             | |||||||   || ||||  || |   | |||||    | ||||  | |  | |

Db    401 CGCTACAACATCACGGGCTGCCTCCTGAGCGCCGACCGCGGGCGGTACGGCGAGGCTCGG 460

Qy    148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAG-- 205
          || || ||  |||| |||| ||| |       | |         |    |  | | |  |
Db    461 GCCGTGCTCTTCCACCACCGCGACCTGGCGCTGCACGGCCGCCAGGGGCTGCCCCGCGGG 520

Qy    206 -------CTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCAC 258
              | | || ||    |||  ||| | |||||||| || || || ||  | |||
Db    521 CCCCCGCCGCGACCCCCGCGGCAGCGGTGGGTGTGGATGAACTTCGAGTCGCCCTCGCAC 580

Qy    259 ACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGAT 318
           | |||     | | ||| | |       | || |||  |||  || | |||| || ||
Db    581 TCCCCCGGGCTGCGGGGCCTGGCCGGGCTCTTCAACTGGACCATGTCGTACCGACGGGAC 640

Qy    319 TCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAA 378
          || ||  ||   ||||| || || | | |     |       |    ||       || |
Db    641 TCGGACGTCTTCGTGCCCTACGGGTACCTCTACGAGCCGCCGTCGCCGCGGCCCTTCGTG 700

Qy    379 GTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCC 438
           ||||  |||||       |||||  ||||||  | || |||||||||   ||||| |||
Db    701 CTGCCTCGCAAAAGCCGGCTGGTGGCCTGGGTCATCAGCAACTGGAACGAGGAGCACGCC 760

Qy    439 AGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 498
           | ||    || |||    |||| |   |   | |    ||| |    ||||||   ||
Db    761 CGCGTGCGCTACTACCGGCAGCTGAAGGAGCACCTGCCCATCGACGTGTACGGG---GCG 817

Qy    499 TTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTAT 558
             ||         |    ||     |   || |    ||  | || || |  || || ||
Db    818 CGCGGGATGGCGCTGCTGGAGGGCAGCGTGGTGAAGACGGTGTCGGCCTACAAGTTCTAC 877

Qy    559 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCT 615
          ||  |||| || || ||    |||| ||| |||||||| || ||||| |   | || ||
Db    878 CTGGCCTTCGAGAACTCCCAGCACACGGACTACATCACCGAGAAGCTGTGGAAGAACGCC 937

Qy    616 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 675
          ||    ||  || | || |||||||| || || ||    ||||  ||||| |||   |
Db    938 TTCGCCGCCAGCGCGGTGCCTGTTGTCCTCGGCCCCCGCAGGGCCAACTACGAGCGCTTC 997

Qy    676 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAG 735
          || || || || || ||||| || ||||| || |    |   ||  |  ||| || |
Db    998 ATCCCCGCCGACTCCTTCATCCACGTGGACGACTTCCCCAGCCCACGGCTGCTGGCCACC 1057

Qy    736 TATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          || ||||||    |||| ||||||||       |||  || || ||   |||| ||||
Db   1058 TACCTGAAGTTCCTCGATAAAAACAAGCCCAGCTACAGGAGGTATTTCGCCTGGCGGAAC 1117

Qy    796 GATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATCAT 855
           | |     ||  |  |  |    || |||||    |||   ||   || ||||||
Db   1118 AAGTATGAAGTCCACGTCACGTCTTTCTGGGATGAGCATTACTGCAAGGTTTGCGAGGCC 1177

Qy    856 GTGAAAAGG 864
          ||||  | |
Db   1178 GTGAGGACG 1186


Search completed: October  6, 2002, 04:44:21
Job time : 1255.59 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:38:25 ; Search time 102.682 Seconds
(without alignments)
15299.444 Million cell updates/sec

Title: US-09-744-748-5_COPY_460_1374
Perfect score: 915
Sequence: 1 accaaaactgattattttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	915	100.0	2676	21	AAZ92647	Human alpha-1,3-fu
2	915	100.0	2822	21	AAZ92646	Human alpha-1,3-fu
3	791.8	86.5	2036	21	AAZ92672	Murine alpha-1,3-f
4	791.8	86.5	2170	21	AAZ92645	Murine alpha-1,3-f
5	127.4	13.9	1814	18	AAT59505	Murine myeloid-lin
6	127	13.9	3594	18	AAT85219	Mouse alpha-fucosy
7	114.4	12.5	1701	15	AAQ77732	Human alpha-1,3-fu
8	114.4	12.5	1701	21	AAF21136	Human low adenosin
9	114.4	12.5	1701	21	AAA35014	Human adenosine re
10	114.4	12.5	6944	21	AAF21137	Human low adenosin
11	114.4	12.5	6944	21	AAA35015	Human adenosine re
12	113.4	12.4	1654	15	AAQ56905	pCDNA1-alpha-(1-3)
13	113.4	12.4	1654	18	AAT61680	Human alpha(1,3)-f
14	112.2	12.3	1086	15	AAQ56912	Lewis blood group
15	112.2	12.3	2042	18	AAT80111	Fucosyltransferase
16	112.2	12.3	2043	12	AAQ13330	Glycosyltransferas
17	112.2	12.3	2043	15	AAQ56906	DNA encoding a gly
18	112.2	12.3	2043	18	AAT76769	Human alpha 1,3/4
19	112.2	12.3	2043	18	AAT61675	Human alpha(1,3/1,
20	111.8	12.2	1126	21	AAF21135	Human low adenosin
21	111.8	12.2	1126	21	AAA35013	Human adenosine re
22	111.8	12.2	1316	15	AAQ56911	DNA encoding a gly
23	111.8	12.2	1316	18	AAT61679	Human alpha(1,3)-f
24	103.6	11.3	795	17	AAT13798	Alpha-(1,3/1/4)-fu
25	102.4	11.2	2175	11	AAQ06691	CDX, a MILA involv
26	100.8	11.0	1256	21	AAF21134	Human low adenosin
27	100.8	11.0	1256	21	AAA35012	Human adenosine re
28	100.8	11.0	1400	13	AAQ31436	Encodes a HeLa cel
29	100.8	11.0	1488	15	AAQ56910	DNA encoding a gly
30	100.8	11.0	2175	12	AAQ14382	Clone 7.2 encoding
31	100.8	11.0	2175	24	AAS17082	Human cDNA clone 7
32	100.8	11.0	2861	12	AAQ14383	Clone 1 encoding 1
33	100.8	11.0	2861	21	AAF21133	Human low adenosin
34	100.8	11.0	2861	21	AAA35011	Human adenosine re
35	100.8	11.0	2861	24	AAS17083	Human cDNA clone 1
36	100.8	11.0	3647	12	AAQ13333	GDP-Fuc:beta-D-Gal
37	100.8	11.0	3647	15	AAQ56909	DNA encoding a gly
38	100.8	11.0	3647	18	AAT61678	Human alpha(1,3)-f
39	97.6	10.7	2134	18	AAT59506	Human myeloid deri
40	66.6	7.3	1578	23	ABL11867	Drosophila melanog
c 41	66.6	7.3	4229	23	ABL11866	Drosophila melanog
42	52.2	5.7	1353	21	AAZ51685	C. elegans alpha-1
43	44.4	4.9	676	22	AAS41344	cDNA encoding nove
44	44.4	4.9	1479	24	ABA05334	Human fucosyltrans
45	44.4	4.9	2557	24	ABA05333	Human fucosyltrans

ALIGNMENTS

RESULT 1
AAZ92647
ID AAZ92647 standard; cDNA; 2676 BP.
XX
AC AAZ92647;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:5.
XX
KW Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW brain; kidney; recombinant expression; transgenic animal; knockout
KW animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW treatment; cancer; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 295..1374
FT /*tag= a
FT /product= "Human alpha-1,3-fucosyltransferase"
XX
PN WO200006708-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-JP04092.
XX
PR 29-JUL-1998; 98JP-0213823.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

```
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 143-150; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2676 BP; 819 A; 527 C; 476 G; 854 T; 0 other;

  Query Match          100.0%;  Score 915;  DB 21;  Length 2676;
  Best Local Similarity  100.0%;  Pred. No. 7.8e-255;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 519

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 579

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 639

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 699

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 759

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 819

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 879

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 939

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 999

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 1059

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 1119

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 1179

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1180 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 1239

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1240 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 1299

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1300 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1359

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    1360 TGGTTTTGGAATTAA 1374


RESULT 2
AAZ92646
ID   AAZ92646 standard; cDNA; 2822 BP.
XX
AC   AAZ92646;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA, SEQ ID NO:4.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1289..2368
FT                   /*tag=  a
FT                   /product= "Human alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80996.
XX
```

```
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 134-142; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequences
CC   AAZ92646 and AAZ92647 represent cDNAs encoding human
CC   alpha-1,3-fucosyltransferase (AAY80996).
XX
SQ   Sequence 2822 BP; 918 A; 485 C; 477 G; 942 T; 0 other;

  Query Match           100.0%;  Score 915;  DB 21;  Length 2822;
  Best Local Similarity  100.0%;  Pred. No. 7.9e-255;
  Matches  915;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1454 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 1513

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1514 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 1573

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1574 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 1633

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1634 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 1693

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1694 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 1753

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1754 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 1813

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1814 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 1873

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1874 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 1933

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1934 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 1993

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1994 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 2053

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2054 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 2113

Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2114 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 2173

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2174 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 2233

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2234 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 2293

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2294 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 2353

Qy     901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db    2354 TGGTTTTGGAATTAA 2368


RESULT 3
AAZ92672
ID   AAZ92672 standard; cDNA; 2036 BP.
XX
AC   AAZ92672;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase gene exon 1 DNA, SEQ ID NO:30.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Example 8; Page 162-164; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
```

```
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2036 BP; 666 A; 365 C; 348 G; 657 T; 0 other;

  Query Match             86.5%;  Score 791.8;  DB 21;  Length 2036;
  Best Local Similarity   91.6%;  Pred. No. 3.6e-219;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          || |||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db     174 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 233

Qy      61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
           |||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db     234 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 293

Qy     121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
           ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db     294 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 353

Qy     181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
           |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db     354 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 413

Qy     241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
           || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db     414 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 473

Qy     301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
           || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     474 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 533

Qy     361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
           |||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db     534 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 593

Qy     421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
           |||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db     594 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 653

Qy     481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
           || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db     654 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 713

Qy     541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
           ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db     714 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 773

Qy     601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
           ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db     774 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 833
```

```
Qy     661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
           |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Db     834 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 893

Qy     721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
           |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db     894 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 953

Qy     781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
           |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db     954 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1013

Qy     841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
           |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1014 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1073

Qy     901 TGGTTTTGGAATTAA 915
           ||||||||||||||||
Db    1074 TGGTTTTGGAATTAA 1088
```

```
RESULT 4
AAZ92645
ID   AAZ92645 standard; cDNA; 2170 BP.
XX
AC   AAZ92645;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Murine alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; fucose; glycosylation; Lewis epitope;
KW   brain; kidney; recombinant expression; transgenic animal; knockout
KW   animal; FUC-TIV; drug screening; inhibitor; potentiator; diagnosis;
KW   treatment; cancer; murine; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1194
FT                   /*tag= a
FT                   /product= "Murine alpha-1,3-fucosyltransferase"
XX
PN   WO200006708-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-JP04092.
XX
PR   29-JUL-1998;   98JP-0213823.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Kudo T,  Sasaki K;
XX
DR   WPI; 2000-183120/16.
DR   P-PSDB; AAY80995.
XX
PT   Alpha 1,3-fucosyltransferase generating Lewis x but not sialyl-Lewis x
PT   epitope and an antibody recognizing it useful for diagnosis of brain
PT   and kidney disease and cancer.  -
XX
PS   Claim 4; Page 127-134; 172pp; Japanese.
XX
CC   The invention relates to a novel alpha-1,3-fucosyltransferase which
CC   transfers a fucose moiety to galactosyl-beta-1-4-N-acetylglucosamine
CC   (generating the Lewis x or y epitope). It does not transfer a fucose
CC   moiety to alpha-2,3-sialyl-galactosyl-beta-1-4-N-acetylglucosamine
CC   and therefore does not generate the  sialyl-Lewis x epitope. The
CC   invention also relates to DNA sequences encoding alpha-1,3-
CC   fucosyltransferase and expression vectors and host cells comprising
CC   these DNA sequences. The invention additionally encompasses the
```

```
CC   preparation of alpha-1,3-fucosyltransferase via the culture of
CC   transformed cells or by expression of the protein in a transgenic animal;
CC   antibodies which recognise alpha-1,3-fucosyltransferase; methods for
CC   screening potential inhibitors or potentiators of
CC   alpha-1,3-fucosyltransferase activity or expression; the preparation of
CC   compounds having fucose-containing sugar chains by use of the protein;
CC   and knockout non-human animals lacking alpha-1,3-fucosyltransferase.
CC   Alpha-1,3-fucosyltransferase has a similar substrate range to the known
CC   FUC-TIV and is expressed mainly in brain and kidney tissues.
CC   Alpha-1,3-fucosyltransferase, nucleotides which encode it, antibodies,
CC   potentiators and inhibitors may be used for the treatment and diagnosis
CC   of diseases of the brain and kidney, and of cancers. They may be used for
CC   the identification of substances which affect the activity or expression
CC   of alpha-1,3-fucosyltransferase; such substances may be used
CC   therapeutically. The knockout animals can be used to study the mechanisms
CC   of action and expression of alpha-1,3- fucosyltransferase. Sequence
CC   AAZ92645 represents cDNA encoding murine alpha-1,3-fucosyltransferase
CC   (AAY80995), and sequences AAZ92670-Z92673 are murine
CC   alpha-1,3-fucosyltransferase exon 1 sequences obtained by 5' RACE (rapid
CC   amplification of cDNA ends).
XX
SQ   Sequence 2170 BP; 715 A; 395 C; 368 G; 692 T; 0 other;

  Query Match            86.5%;  Score 791.8;  DB 21;  Length 2170;
  Best Local Similarity  91.6%;  Pred. No. 3.7e-219;
  Matches  838;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy      1 ACCAAAACTGATTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAG 60
          || ||||||||||||||||| |||||||| |||||||| ||||| ||||||||||||||||
Db    280 ACAAAAACTGATTATTTTAACGAAACTACCATTCTGGTTTGGGTATGGCCATTTGGGCAG 339

Qy     61 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACG 120
          |||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||
Db    340 ACCTTTGACCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACAACA 399

Qy    121 GACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGG 180
          ||||| ||| |||||||||||||||||| || ||||| || ||  |||||||||| |||
Db    400 GACCGCTCATTGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGCTGG 459

Qy    181 GATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAAT 240
          |||||||| || |||||||||| || |||||||||| ||||||||||||||||||||||
Db    460 GATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGATTTGGATGAAT 519

Qy    241 TTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACT 300
          || || ||||| |||||||| |||||||||||||||||||| |||||||| |||||||||
Db    520 TTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTGACT 579

Qy    301 CTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACA 360
          || ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    580 CTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCACA 639

Qy    361 AATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAAC 420
          |||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||
Db    640 AATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTAAC 699

Qy    421 TGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATC 480
          |||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||||
Db    700 TGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAATC 759

Qy    481 CATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATA 540
          || ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||||
Db    760 CACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCATA 819

Qy    541 TCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAA 600
          ||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |||
Db    820 TCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAGAA 879

Qy    601 AAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAA 660
          ||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||||
Db    880 AAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGAA 939

Qy    661 AACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCC 720
          |||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Db    940 AACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCCC 999

Qy    721 AGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTAC 780
          |||||| ||||||| |||||||||||||| |||||||||||||||| ||||||||||||
Db   1000 AGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTAC 1059

Qy    781 TTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGT 840
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db   1060 TTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATGC 1119

Qy    841 TTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 900
            |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1120 CTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAA 1179

Qy    901 TGGTTTTGGAATTAA 915
          |||||||||||||||
Db   1180 TGGTTTTGGAATTAA 1194


RESULT 5
AAT59505
ID   AAT59505 standard; cDNA; 1814 BP.
XX
AC   AAT59505;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Murine myeloid-lineage alpha-(1,3)-fucosyltransferase cDNA.
XX
KW   Alpha-(1,3)-fucosyltransferase; fucosylation; antibody; IgG; IgM;
KW   septic shock; septicaemia; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             325..1353
FT                   /*tag=  a
XX
PN   WO9640881-A1.
XX
PD   19-DEC-1996.
XX
PF   08-MAY-1996;   96WO-US06427.
XX
PR   07-JUN-1995;   95US-0483151.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Holgersson J,  Seed B;
XX
DR   WPI; 1997-108639/10.
DR   P-PSDB; AAW11820.
XX
PT   New murine alpha-(1,3)-fucosyltransferase - for fucosylating an
PT   antibody to protect mammals against e.g. septic shock or septicaemia
XX
PS   Claim 6; Fig 3; 58pp; English.
XX
CC   A cDNA clone (AAT59505) codes for murine myeloid lineage alpha-
CC   (1,3)-fucosyltransferase (AAW11820), an enzyme which has a relatively
CC   strict substrate requirement for sialylated N-acetyllactosamine,
CC   which can account for the presence of the sialyl-Le-w epitope on
CC   murine cells, and which is more effective than Fuc-TIV in support
CC   of E-selectin-mediated COS cell adhesion.  It was isolated from a
CC   murine myeloid 32D c13 cell cDNA library by identification of a
CC   clone capable of directing the expression of sialyl-Le-x
CC   determinants.  Transformed host cells (e.g. 32D c13 or human 293
CC   cells) expressing the alpha-(1,3)-fucosyltransferase, and pref.
CC   another fucosyltransferase such as human Fuc-TIV (see also AAT59506),
CC   can be used to fucosylate an antibody (e.g. IgG or IgM) for use in
CC   protecting an animal against an adverse immune reaction, esp. septic
CC   shock or septicaemia.
XX
```

```
SQ   Sequence 1814 BP; 360 A; 523 C; 518 G; 413 T; 0 other;

  Query Match             13.9%;  Score 127.4;  DB 18;  Length 1814;
  Best Local Similarity   52.1%;  Pred. No. 1.3e-26;
  Matches  334;  Conservative    0;  Mismatches  301;  Indels    6;  Gaps    2;

Qy     104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
           | |||| ||| |   | ||||   |||     ||   |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |          |   |     |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
           ||| | |||       ||||||| || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGCTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           | | || |||  |   ||||  || || ||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
           ||||  |     | |    ||   |   |    | | ||   ||||   |  |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           ||||||  | || || |   |      |||     | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
           |      |  |  |  ||   | ||| |  ||    |||        |    | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
           | || | || ||  |  |     |     || || ||  ||||||| || |||   ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           ||| ||||||||| || ||| | |    ||||||  |   ||||||  ||||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||  || |||  | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTG 741
           |||| || |  | |||| || |||| || ||     |  |||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGCTGTCTTCCTG 1185


RESULT 6
AAT85219
ID   AAT85219 standard; DNA; 3594 BP.
XX
AC   AAT85219;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Mouse alpha-fucosyltransferase Fuc-TVII gene.
XX
KW   Alpha-fucosyltransferase; Fuc-TVII gene; mouse;
KW   sialyl Lewis x determinant; oligosaccharide; antiinflammatory;
KW   inhibitor; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             996..3082
FT                   /*tag=  a
FT                   /note= "includes introns"
FT   exon            465..477
FT                   /*tag=  b
FT                   /number= 1
FT   intron          478..669
FT                   /*tag=  c
FT                   /number= 1
FT   exon            670..1149
FT                   /*tag=  d
FT                   /number= 2
FT                   /codon_start= 996..998
FT   intron          1150..1946
FT                   /*tag=  e
FT                   /number= 2
FT   exon            1947..1959
FT                   /*tag=  f
FT                   /number= 3
FT                   /codon_start= 1947..1949
FT                   /note= "alternative ATG start codon"
FT   intron          1960..2065
FT                   /*tag=  g
FT                   /number=  3
FT   exon            2067..3552
FT                   /*tag=  f
FT                   /number= 4
FT                   /codon_start= 2126..2128
FT                   /note= "alternative ATG start codon"
XX
PN   WO9732889-A1.
XX
PD   12-SEP-1997.
XX
PF   07-MAR-1997;   97WO-US03184.
XX
PR   08-MAR-1996;   96US-0613098.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Gersten KM,  Lowe JB,  Natsuka S;
XX
DR   WPI; 1997-470535/43.
DR   P-PSDB; AAW27138 and AAW26671.
XX
PT   Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.
PT   screening for inhibitors used as antiinflammatory agents
XX
PS   Claim 1; Fig 2; 91pp; English.
XX
CC   This DNA sequence comprises the mouse Fuc-TVIII gene encoding
CC   alpha-fucosyltransferase.  It was isolated from a 3T3 cell genomic
CC   library by screening with a probe from the catalytic domain of the
CC   human Lewis alpha(1,3/1,4)fucosyltransferase (Fuc-TIII).
CC   Examination of the mouse Fuc-TVIII locus identified 3 Met codons
CC   that may initiate translation of alpha(1,3) fucosyltransferases
CC   with different cytosolic domains encoded by exons 2 and/or 3, but
CC   with identical Golgi-localised catalytic domains encoded by exon 4
CC   (see AAW27138 and AAW26671).  The proteins and nucleic acids can
CC   be used to construct animal cell lines able to post-translationally
CC   modify oligosaccharides on cell surface, intracellular or secreted
CC   proteins and lipids, to isolate reagents for efficient enzymatic
CC   production of oligosaccharides, to generate antibodies to
CC   glycosyltransferases useful as diagnostic reagents, to screen for
CC   fucosyltransferase inhibitors and inactivators, especially those
CC   that act as antiinflammatory agents, for genotyping individuals at
CC   the fucosyltransferase locus and for in vitro synthesis of sialyl
CC   Lewis X tetrasaccharide.
XX
SQ   Sequence 3594 BP; 731 A; 969 C; 1067 G; 827 T; 0 other;

  Query Match             13.9%;  Score 127;  DB 18;  Length 3594;
  Best Local Similarity   52.3%;  Pred. No. 2.4e-26;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
           | |||| ||| |   | ||||   |||     ||   |  |||| ||  |  |||| |
Db    2277 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 2336
```

```
Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |           |   |     |||| | | |    ||||||| |   |||
Db    2337 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 2396

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
           ||| | |||       |||||||| || | | | || |          | || |   |
Db    2397 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 2456

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           | | || |||  |   |||| || || ||||||||||||||||  || || || ||
Db    2457 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 2516

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
           ||||  |     | |    ||   |   |    | | ||   ||||  |   |||
Db    2517 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 2573

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           ||||||  | || || |   |       |||     | |  |||   |||    ||||
Db    2574 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 2633

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
           |       |  |  |  ||  | ||| |  ||    |||       |    | |||
Db    2634 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 2693

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
           | || | || ||  |  |     |      || || ||  |||||| || |||   ||
Db    2694 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 2753

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           |||  |||||||| || ||| | |    ||||||   |   ||||||  |||||| || | |
Db    2754 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 2813

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||  || |||  | ||| |||  || | | ||| ||||| | ||  | || |
Db    2814 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 2873

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           |||| || |  | |||| || |||| || ||
Db    2874 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 2904
```

```
RESULT 7
AAQ77732
ID   AAQ77732 standard; cDNA to mRNA; 1701 BP.
XX
AC   AAQ77732;
XX
DT   22-JUN-1995  (first entry)
XX
DE   Human alpha-1,3-fucosyltransferase cDNA.
XX
KW   Alpha-1,3-fucosyltransferase; sialylated Lewis blood group antigen;
KW   Lewis x; Lewis a; glycosylation; glycoprotein; glycolipid; ds.
XX
OS   Homo sapiens (THP-1 cells).
XX
FH   Key             Location/Qualifiers
FT   CDS             123..1151
FT                   /*tag= a
FT                   /product= alpha-1,3-fucosyltransferase
XX
PN   WO9423021-A.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994;   94WO-JP00496.
XX
PR   29-MAR-1993;   93JP-0069016.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hanai N,  Kurata K,  Nishi T,  Sasaki K;
XX
DR   WPI; 1994-333183/41.
DR   P-PSDB; AAR63215.
XX
PT   Alpha-1,3-Fucosyl-transferase DNA and protein - useful for prodn.
PT   of active sugar chains, e.g. sialylated Lewis X
XX
PS   Claim 3; Page 99-102; 155pp; Japanese.
XX
CC   The sequence coding for alpha-1,3-fucosyltransferase was isolated
CC   from a human THP-1 cell cDNA library. The enzyme catalyses the
CC   addition of fucose via a 1,3-linkage to N-acetylglucosamine
CC   contained in glycoproteins and glycolipids, e.g. for sialylation of
CC   Lewis blood group antigens.
XX
SQ   Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

  Query Match             12.5%;  Score 114.4;  DB 15;  Length 1701;
  Best Local Similarity   50.8%;  Pred. No. 7.6e-23;
  Matches  327;  Conservative    0;  Mismatches  311;  Indels    6;  Gaps    2;
```

```
Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | ||||| || |   |  ||||    |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |           | | |  |  | || | |       ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |    |||   ||| | |||        |||| || || |  |||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |      | | || ||| |   ||||  ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           |||||  ||||||  | ||  |      |      ||||    |       |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
              |   |||    |||||| || || ||||   |      |||     | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||           |      |  ||  ||  || |  ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |    |  |       || | || ||| |  |    |      || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           |||| || || || ||||| ||  |  |||  | |||||||    |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

```
RESULT 8
AAF21136
ID   AAF21136 standard; DNA; 1701 BP.
XX
AC   AAF21136;
XX
DT   14-MAR-2001  (first entry)
```

XX
DE Human low adenosine antisense oligonucleotide related sequence #2703.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 1701 BP; 290 A; 527 C; 589 G; 295 T; 0 other;

Query Match 12.5%; Score 114.4; DB 21; Length 1701;
Best Local Similarity 50.8%; Pred. No. 7.6e-23;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||   | |||||  || |  |  ||||    |||     ||   || | ||
Db     333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |           | | |  |  | || | |       ||
Db     393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |     |||   ||| | |||       |||| || || |  |||| | |
Db     453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |       | | || ||| |   |||| ||||| || || || || |||
Db     513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           ||||| ||||||  | ||  |   |     ||||   |       |||||  |||
Db     573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
              |  |||    |||||| || || ||||   |     |||     | | || |
Db     630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||             |    | ||  ||  || | ||   ||||
Db     690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
             |   | |      || | || ||| |  |    |     || || || ||||||||
Db     750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db     810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
           |||| || || || ||||| ||  |  |||  | |||||||   |   | || || ||
Db     870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | ||| | |||||||| || | |  |||  || | ||||| ||
Db     930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 9
AAA35014
ID AAA35014 standard; DNA; 1701 BP.
XX
AC AAA35014;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2703.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.

XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 957-958; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 1701 BP; 292 A; 524 C; 586 G; 299 T; 0 other;

Query Match 12.5%; Score 114.4; DB 21; Length 1701;
Best Local Similarity 50.8%; Pred. No. 7.6e-23;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy     91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
Db    333 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 392

Qy    151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
Db    393 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 452

Qy    211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
Db    453 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 512

Qy    271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
Db    513 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 572

Qy    331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
Db    573 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 629

Qy    391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
Db    630 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 689

Qy    451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
Db    690 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 749

Qy    511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
Db    750 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 809

Qy    571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
Db    810 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 869

Qy    628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
Db    870 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 929

Qy    688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
Db    930 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 973
```

RESULT 10
AAF21137
ID AAF21137 standard; DNA; 6944 BP.
XX
AC AAF21137;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2704.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1034-1036; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,

CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

Query Match 12.5%; Score 114.4; DB 21; Length 6944;
Best Local Similarity 50.8%; Pred. No. 1.4e-22;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | |||||  || |   |  ||||    |||     ||   || | ||
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           ||  |  |||| ||||| ||  |          | | |  |  | || | |       ||
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || |     |||   ||| | |||       |||| || || |  ||||| | |
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           ||   | |      | | || |||  |   ||||  ||||| || || || || |||
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           |||||  ||||||  | ||  |      |      ||||     |      |||||  |||
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
              |  |||    |||||| || || ||||   |      |||     | | || |
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           |||    ||||           |      |  ||  ||  ||  |  ||   ||||
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
            |   |  |      || | || ||| |  |    |     || || || ||||||||
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
           || ||    |||   || ||||| ||||| ||  | |    ||| ||  |  ||||||||
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
            |||| || || || ||||| ||  |  |||  | |||||||    |    | || || ||
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           | |||  | |||||||| || | |  |||  || | ||||| ||
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

RESULT 11
AAA35015
ID AAA35015 standard; DNA; 6944 BP.

XX
AC AAA35015;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2704.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 958-960; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 6944 BP; 1218 A; 2193 C; 2223 G; 1310 T; 0 other;

Query Match 12.5%; Score 114.4; DB 21; Length 6944;
Best Local Similarity 50.8%; Pred. No. 1.4e-22;
Matches 327; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

```
Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | |  ||||    | |||||  || |   |  ||||    |||     ||   || | ||
```

```
Db    5576 TACGGCATCGCCCGCTGCCACCTGAGTGCCAACCGAAGCCTGCTGGCCAGCGCCGACGCC 5635

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
Db    5636 GTGGTCTTCCACCACCGCGAGCTGCAGACCCGGCGGTCCCACCTGCCCCTGGCCCAGCGG 5695

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
Db    5696 CCGCGAGGGCAGCCCTGGGTGTGGGCCTCCATGGAGTCTCCTAGCCACACCCACGGCCTC 5755

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
Db    5756 AGCCACCTCCGAGGCATCTTCAACTGGGTGCTGAGCTACCGGCGCGACTCGGACATCTTT 5815

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
Db    5816 GTGCCCTATGGCCGCCTGGAGCCCCACTGGGGGCCCT---CGCCACCGCTGCCAGCCAAG 5872

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
Db    5873 AGCAGGGTGGCCGCCTGGGTGGTCAGCAACTTCCAGGAGCGGCAGCTGCGTGCCAGGCTG 5932

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
Db    5933 TACCGGCAGCTGGCGCCTCATCTGCGGGTGGATGTCTTTGGCCGTGCCAATGGACGGCCA 5992

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
Db    5993 CTGTGCGCCAGCTGCCTGGTGCCCACCGTGGCCCAGTACCGCTTCTACCTGTCCTTTGAG 6052

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
Db    6053 AACTCTCAGCACCGCGACTACATTACGGAGAAATTCTGGCGCAACGCACTGGTGGCTGGC 6112

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687
Db    6113 ACTGTGCCAGTGGTGCTGGGGCCCCCACGGGCCACCTATGAGGCCTTCGTGCCGGCTGAC 6172

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
Db    6173 GCCTTCGTGCATGTGGATGACTTTGGCTCAGCCCGAGAGCTGGC 6216
```

```
RESULT 12
AAQ56905
ID   AAQ56905 standard; DNA; 1654 BP.
XX
AC   AAQ56905;
XX
DT   26-JUL-1994  (first entry)
XX
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1208
FT                   /*tag=  a
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
XX
DR   WPI; 1994-048874/06.
DR   P-PSDB; AAR45933.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Claim 3; Fig 7; 249pp; English.
XX
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56906-12.
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;
```

```
  Query Match             12.4%;  Score 113.4;  DB 15;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.5e-22;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
```

```
Db   1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy    793 AAG 795
          ||
Db   1095 GAG 1097


RESULT 13
AAT61680
ID   AAT61680 standard; DNA; 1654 BP.
XX
AC   AAT61680;
XX
DT   19-JUN-1997  (first entry)
XX
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
XX
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW   glycosylation; oligosaccharide; blood group; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..1208
FT                   /*tag= a
XX
PN   WO9709421-A1.
XX
PD   13-MAR-1997.
XX
PF   06-SEP-1996;   96WO-US13816.
XX
PR   08-SEP-1995;   95US-0525058.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Legault DJ,  Lowe JB;
XX
DR   WPI; 1997-192897/17.
DR   P-PSDB; AAW13643.
XX
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
XX
PS   Example 6; Page 291-292; 329pp; English.
XX
CC   A DNA clone (AAT61680) codes for an enzyme (AAW13643) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC   precursors but not type I glycoconjugates.  It was identified in a
CC   human lambda phage genomic DNA library using a probe obtd. from
CC   Lewis enzyme cDNA (see also AAT61675).  The Fuc-TVI DNA can be used
CC   to construct animal cell lines with specific capabilities with
CC   respect to post-translational modification of the oligosaccharides
CC   of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC   use in oligosaccharide prodn., to genotype individuals at this
CC   fucosyltransferase locus, and in the prodn. of chimeric
CC   fucosyltransferases (see also AAW14512-32).
XX
SQ   Sequence 1654 BP; 298 A; 557 C; 459 G; 340 T; 0 other;

  Query Match             12.4%;  Score 113.4;  DB 18;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.5e-22;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  ||||||  |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
           |       ||      |  ||| |  ||||  | |||||     ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |    .|||  |||||| ||| | |    |||| || |||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||          | | | |     |  | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||     |  |||||  |||||  |||   ||||||   || |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           |||||  ||    || ||| |    ||              |  |  |  |  |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           |      |      |     |   ||   ||||   ||  | ||   |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| |  ||  | ||||| || ||||| || |    |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
           ||| ||| ||||  | |||||  || |      ||||| || ||||||   |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           ||
Db    1095 GAG 1097


RESULT 14
AAQ56912
ID   AAQ56912 standard; DNA; 1086 BP.
XX
AC   AAQ56912;
XX
DT   26-JUL-1994  (first entry)
XX
DE   Lewis blood group fucosyltransferase DNA.
XX
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
XX
PN   WO9402616-A.
XX
PD   03-FEB-1994.
XX
PF   20-JUL-1993;   93WO-US06703.
XX
PR   20-JUL-1992;   92US-0914281.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Lowe JB;
```

```
XX
DR   WPI; 1994-048874/06.
XX
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
XX
PS   Disclosure; Fig 6; 249pp; English.
XX
CC   The sequence is that encoding a Lewis blood group fucosyl
CC   transferase (Fuc-TIII).  The enzyme produced by the DNA may be
CC   non glycosylated.  This prevents premature loss of enzyme activity.
CC   It can also be used in in vitro reactions to modify cell surface
CC   oligosaccharide mols. e.g. blood group determinants.
CC   See also AAQ56905-11.
XX
SQ   Sequence 1086 BP; 203 A; 387 C; 298 G; 198 T; 0 other;

  Query Match             12.3%;  Score 112.2;  DB 15;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.7e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
           |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db     271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy     166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
           | || ||||  |    |   |  | ||    | || |       | |||||  |     |||
Db     331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy     223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
           |||||  |||  | ||  |||||  ||||  || ||   |  ||   |   | | | ||
Db     391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy     283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
           |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db     451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy     343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
           |   ||  |    | |    |     |          |-| |  |  |  |   |||
Db     511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy     394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db     571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy     454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
           |    ||           |  |  |  |    |||||  |   |                |
Db     631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
           || |  |  |  ||||     ||  | ||     |  || || || ||  |||| || ||
Db     691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
           ||  | |||   || ||||||||| || ||||| |      || ||  |   |||  |  |
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
           || || || || ||||| ||    ||    |||||  ||||  |   | ||| | ||  |
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
           |||||  || ||||| || |    |    ||||  || || ||   ||| ||| ||||  |
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
           |||||  || |       ||||| || ||||||  |||| || ||
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 15
AAT80111
ID   AAT80111 standard; cDNA; 2042 BP.
XX
AC   AAT80111;
XX
DT   01-DEC-1997  (first entry)
XX
DE   Fucosyltransferase III coding sequence.
XX
KW   P-selectin glycoprotein ligand-1; PSGL-1; human; mucin-like protein;
KW   calcium-dependent carbohydrate binding protein; platelet; endothelium;
KW   thrombin; leukocyte; sulphated glycosylated peptide; O-glycan; therapy;
KW   inflammation; ischaemia; reperfusion; bacterial sepsis; atherosclerosis;
KW   disseminated intravascular coagulation; rheumatoid arthritis; antibody;
KW   adult respiratory distress syndrome; tumour metastasis;
KW   fucosyltransferase III; acetylglucosaminyltransferase;
KW   autoimmune disorder; inflammatory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9706176-A2.
XX
PD   20-FEB-1997.
XX
PF   02-AUG-1996;   96WO-US12820.
XX
PR   17-MAY-1996;   96US-0649802.
PR   03-AUG-1995;   95US-0510920.
PR   15-MAY-1996;   96US-0017794.
XX
PA   (OKLA ) UNIV OKLAHOMA STATE.
XX
PI   Cummings RD,  Mcever RP,  Moore KL;
XX
DR   WPI; 1997-154206/14.
XX
PT   New O-glycan(s) derived from P-selectin glyco:protein ligand
PT   (PSGL)-1 - used for inhibiting binding of PSGL-1 to selectins, e.g.
PT   for treating inflammatory or autoimmune disorders or tumours
XX
PS   Disclosure; Page 80-81; 96pp; English.
XX
CC   This sequence represents the coding sequence for the human
CC   fucosyltransferase III. The encoded protein, and the core 2 beta1-6-N-
CC   acetylglucosaminyltransferase encoded by AAT80112 are needed for the
CC   expression of the human P-selectin glycoprotein ligand-1 (PSGL-1) in
CC   mammalian cells. P-selectin is a calcium-dependent carbohydrate binding
CC   protein expressed on the surfaces of activated platelets and endothelium
CC   in response to thrombin and other agonists. PSGL-1 (see AAW26174) is a
CC   high affinity P-selectin ligand produced by leukocytes. Binding of
CC   P-selectin to PSGL-1 is calcium ion dependent and is abolished by
CC   treatment of the ligand with sialidase. PSGL is a homodimer, with a
CC   highly extended extracellular domain, which is a feature of mucin-like
CC   proteins. PSGL-1 is targeted by the sulphated glycosylated peptides of
CC   the invention (see AAW26175-W26188). PSGL-1 is also targeted by the
CC   O-glycans, and O-glycanated products of the invention. The products are
CC   used for inhibiting the binding of P-selectin and other selectins to
CC   PSGL-1. They can be used for the treatment of inflammation, injury
CC   resulting from ischaemia and reperfusion, bacterial sepsis and
CC   disseminated intravascular coagulation, adult respiratory distress
CC   syndrome, tumour metastasis, rheumatoid arthritis, atherosclerosis and
CC   other autoimmune or inflammatory disorders. The products can also be used
CC   in the production of antibodies and in diagnostic applications.
XX
SQ   Sequence 2042 BP; 413 A; 646 C; 575 G; 408 T; 0 other;

  Query Match             12.3%;  Score 112.2;  DB 18;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 3.6e-22;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
           |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402
```

```
Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || |||| |   |   | | ||    | || |      | ||||| |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          ||||| ||| | || ||||| |||| || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |    |  |    |     |           | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||          |  |  |  |    |||||  |  |              |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || |||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||    ||    ||||| ||||  |   | ||| | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |   |    |||| || || ||   ||| ||| ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |       ||||| || ||||||  |||| || ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 5, 2002, 23:16:20
Job time : 144.682 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:40:30 ; Search time 766.105 Seconds
(without alignments)
16120.141 Million cell updates/sec

Title: US-09-744-748-5_COPY_460_1374
Perfect score: 915
Sequence: 1 accaaaactgattattttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	573.8	62.7	697	9	BB397507	BB397507 BB397507
2	492.8	53.9	607	10	BI102760	BI102760 602888338
3	422	46.1	716	9	AV725071	AV725071 AV725071
4	324.8	35.5	649	9	AV338075	AV338075 AV338075
c 5	307.4	33.6	579	10	BM021081	BM021081 ie75f03.x
c 6	286	31.3	1101	12	CNS059ST	AL327638 Tetraodon
c 7	184.2	20.1	570	10	BM082484	BM082484 fu25e08.x
c 8	177	19.3	1100	12	CNS0579Z	AL324368 Tetraodon
c 9	170.2	18.6	1016	12	CNS0429P	AL271222 Tetraodon
10	161.4	17.6	844	12	CNS01V8V	AL168808 Tetraodon
c 11	157.4	17.2	887	12	CNS02Q0S	AL208693 Tetraodon
12	155.4	17.0	651	10	BJ036855	BJ036855 BJ036855
c 13	151.6	16.6	666	10	BJ050300	BJ050300 BJ050300
c 14	151.2	16.5	711	12	CNS034XF	AL228012 Tetraodon
c 15	151	16.5	501	12	B41073	B41073 HS-1052-B2-
16	146.6	16.0	640	10	BJ073227	BJ073227 BJ073227
17	145.2	15.9	648	9	BB595164	BB595164 BB595164
18	136.2	14.9	199	9	BB398103	BB398103 BB398103
19	133	14.5	625	9	BB649967	BB649967 BB649967
20	123.8	13.5	767	9	BB640911	BB640911 BB640911
c 21	123.4	13.5	674	10	BE605503	BE605503 fl10h04.x
22	120.6	13.2	971	12	CNS047IV	AL278032 Tetraodon
c 23	120.4	13.2	615	10	BI326811	BI326811 STR00078
c 24	117.8	12.9	1023	12	CNS02ANG	AL188773 Tetraodon
25	102.8	11.2	487	10	BE863396	BE863396 UI-M-BH0-
c 26	100.2	11.0	702	10	BJ043578	BJ043578 BJ043578
27	93.8	10.3	863	12	CNS02ROF	AL210840 Tetraodon
28	90.4	9.9	720	10	BF180380	BF180380 601804564
29	90	9.8	551	10	BJ035701	BJ035701 BJ035701
30	87.4	9.6	601	10	BJ036679	BJ036679 BJ036679
31	87.4	9.6	605	10	BJ028804	BJ028804 BJ028804
c 32	87	9.5	1056	12	CNS01VL7	AL169252 Tetraodon
33	86.8	9.5	304	9	AU081478	AU081478 AU081478
34	84.2	9.2	638	10	BJ035697	BJ035697 BJ035697
35	83.6	9.1	1057	12	CNS03ADC	AL235065 Tetraodon
c 36	83	9.1	566	10	BJ090565	BJ090565 BJ090565
37	80.2	8.8	628	9	BB651409	BB651409 BB651409
c 38	79.4	8.7	776	12	CNS03RIO	AL257289 Tetraodon
39	78	8.5	568	10	BM021317	BM021317 ie75f03.y
40	78	8.5	838	10	BG432556	BG432556 602500710
41	76.4	8.3	655	10	BG538740	BG538740 602566977
c 42	75	8.2	878	12	CNS02V1D	AL215194 Tetraodon
43	71.6	7.8	590	10	BM082755	BM082755 fu25e08.y
44	70.4	7.7	582	9	AW233458	AW233458 fj32h01.y
45	70	7.7	505	10	BJ029303	BJ029303 BJ029303

ALIGNMENTS

RESULT 1
BB397507
LOCUS BB397507 697 bp mRNA linear EST 24-OCT-2001
DEFINITION BB397507 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330004C19 3' similar to AB015426 Mus musculus Fut9 mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION BB397507
VERSION BB397507.2 GI:16411808
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jul 15, 2000 this sequence version replaced gi:9216903.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330004C19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      219 a    138 c    141 g    199 t
ORIGIN
```

Query Match 62.7%; Score 573.8; DB 9; Length 697;
Best Local Similarity 91.4%; Pred. No. 4.4e-148;
Matches 619; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

```
Qy    239 ATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGA 298
          | || || ||||| |||||||| ||| |||||||||||| ||| |||||||| |||||||
Db      4 AATTAGAGTCACCCACTCACACCCCCAAAAAGAGTGGCA-TGAACACTTGTTCAACCTGA 62

Qy    299 CTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCA 358
          |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     63 CTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGCA 122

Qy    359 CAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTA 418
          |||||||||| |||||||||||||||||| ||||| ||||||||||||||||||||||
Db    123 CAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGTA 182

Qy    419 ACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAA 478
          |||||||||||||||||||| |||||||||||||| ||||| ||||| || |||||||
Db    183 ACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAAA 242

Qy    479 TCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCA 538
          |||| ||||| || |||||||| |||||||| || || ||||||||| ||||||| ||||
Db    243 TCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACCA 302

Qy    539 TATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGG 598
          ||||| |||||||||||||||||| |||||||| ||||| ||||| |||||||||| |
Db    303 TATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACAG 362

Qy    599 AAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGG 658
          ||||||| |||||||| ||| ||||||| || |||||||||||| ||||| ||||||||||
Db    363 AAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGG 422

Qy    659 AAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTC 718
          |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||
Db    423 AAAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTC 482

Qy    719 CCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTT 778
          |||||||| ||||||| |||||||||||||| |||||||||||||||||| ||||||||||
Db    483 CCAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTT 542

Qy    779 ACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCAT 838
          |||||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||
Db    543 ACTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCAT 602

Qy    839 GTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 898
          |  |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db    603 GCCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGA 662

Qy    899 AATGGTTTTGGAATTAA 915
          ||||||||||||||||
Db    663 AATGGTTTTGGAATTAA 679
```

```
RESULT 2
BI102760
LOCUS       BI102760                607 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888338F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043609
            5', mRNA sequence.
ACCESSION   BI102760
VERSION     BI102760.1  GI:14553653
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: 1  column: 10
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043609"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      179 a    149 c    121 g    157 t      1 others
ORIGIN
```

Query Match 53.9%; Score 492.8; DB 10; Length 607;
Best Local Similarity 90.1%; Pred. No. 1.2e-125;
Matches 549; Conservative 0; Mismatches 58; Indels 2; Gaps 2;

```
Qy      58 CAGACCTTTGACCTTACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACA 117
           |||||||||||| |||||||||||||||||||||||||| |||||||| ||||||||||||
Db       1 CAGACCTTTGAGCTTACATCCTGCCAAGCAATGTTCAATATCCAAGGGTGCCATCTCACA 60

Qy     118 ACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGT 177
           || ||||| ||| |||||||||||||||||| || ||||| || ||  ||||||||||
Db      61 ACAGACCGCTCA-TGTACAACAAATCCCATGCGGTCCTGATTCACCATAGAGACATCAGC 119

Qy     178 TGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATG 237
           |||||||||||| || |||||||||||| || |||||||||| |||||||||| |||||||
Db     120 TGGGATCTGACTAACTTACCTCAGCAGGCCAGGCCACCCTTTCAGAAATGGA-TTGGATG 178

Qy     238 AATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTG 297
           ||||| || ||||| |||||||| |||||||||||||||||||| |||||||| ||||||
Db     179 AATTTAGAGTCACCCACTCACACCCCCCAAAAGAGTGGCATTGAACACTTGTTCAACCTG 238

Qy     298 ACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGC 357
           ||||| ||||| |||||||||||||||||||||||||||||||||||||||||||| |||
Db     239 ACTCTAACTTATCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTGAGC 298

Qy     358 ACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGT 417
           |||||||||||| |||||||||||||||||||| ||||| ||||||||||||||||||||
Db     299 ACAAATCCCTTTGTGTTTGAAGTGCCAAGCAAGGAGAAGTTGGTGTGCTGGGTTGTGAGT 358

Qy     418 AACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAA 477
           |||||||||||||||||||||| |||||||||||||| ||||| |||||| || ||||||
Db     359 AACTGGAACCCTGAGCATGCCAGGGTCAAGTATTACAACGAGCTCAGCAAGAGTATTGAA 418

Qy     478 ATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACC 537
           ||||| ||||| || |||||||| |||||||| || || |||||||| ||||||| |||
Db     419 ATCCACACCTATGGCCAAGCATTCGGAGAATACGTGAACGATAAAAATCTGATTCCCACC 478

Qy     538 ATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACG 597
           |||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||||||||
Db     479 ATATCTACTTGTAAATTTTATCTTTCATTTGAAAACTCAATTCACAAAGATTACATCACA 538

Qy     598 GAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGG 657
           |||||||| |||||||| ||| ||||||| || ||||||  ||| ||||| |||||||||
Db     539 GAAAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTNGTGTCCTGGGTCCATCTAGG 598

Qy     658 GAAAACTAT 666
           |||||||||
Db     599 GAAAACTAT 607
```

```
RESULT 3
AV725071
LOCUS       AV725071                716 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725071 HTC Homo sapiens cDNA clone HTCBLB02 5', mRNA sequence.
ACCESSION   AV725071
VERSION     AV725071.1  GI:10830091
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .716
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCBLB02"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      219 a    113 c    133 g    249 t      2 others
ORIGIN

  Query Match             46.1%;  Score 422;  DB 9;  Length 716;
  Best Local Similarity   100.0%;  Pred. No. 5e-106;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     494 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAAT 60

Qy     554 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATG 120

Qy     614 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATT 180

Qy     674 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAA 240

Qy     734 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGA 300

Qy     794 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTGCGATC 360

Qy     854 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTTTGGAATT 420

Qy     914 AA 915
           ||
Db     421 AA 422
```

```
RESULT 4
AV338075
LOCUS       AV338075                649 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV338075 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430407H24 3' similar to AB015426 Mus musculus
            mRNA for alpha1,3-fucosyltransferase IX, mRNA sequence.
ACCESSION   AV338075
VERSION     AV338075.2  GI:16396626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
```

COMMENT On Nov 11, 1999 this sequence version replaced gi:6378127.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. ..10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Location/Qualifiers
source 1. .649
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6430407H24"
/clone_lib="RIKEN full-length enriched, adult male olfactory bulb"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 209 a 105 c 117 g 218 t
ORIGIN

Query Match 35.5%; Score 324.8; DB 9; Length 649;
Best Local Similarity 91.5%; Pred. No. 4.1e-79;
Matches 344; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy    540 ATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGA 599
          ||| ||||||| ||| |||||| |||||||| ||||| |||||||||||||||||| ||
Db      2 ATCAACTTGTAATTTTATTCTTTCATTTGAAAACTCAATTCACAAGGATTACATCACAGA 61

Qy    600 AAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGA 659
          |||||| |||||||| ||| ||||||| || |||||||||| ||||| ||||||||||
Db     62 AAAGCTCTACAATGCATTTTTGGCTGGTTCAGTACCTGTTGTCCTGGGTCCATCTAGGGA 121

Qy    660 AAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCC 719
          ||||||||||||||||||||||| ||||||||||||||||||||||||| |||||||||
Db    122 AAACTATGAGAATTATATTCCAGCTGATTCATTCATTCATGTGGAAGATTTTAACTCTCC 181

Qy    720 CAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTA 779
          ||||||| ||||||| |||||||||||||| |||||||||||||||||| ||||||||||
Db    182 CAGTGAGTTAGCAAAATATCTGAAGGAAGTTGACAAAAACAATAAGTTGTACCTTAGTTA 241

Qy    780 CTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATG 839
          |||||||||||| |||||||| |||||||| || ||||| ||||||||||||||||||||
Db    242 CTTTAACTGGAGAAAGGATTTTACTGTAAACCTACCACGGTTTTGGGAATCACATGCATG 301

Qy    840 TTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 899
            |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    302 CCTGGCATGCGACCATGTAAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAA 361

Qy    900 ATGGTTTTGGAATTAA 915
          ||||||||||||||||
Db    362 ATGGTTTTGGAATTAA 377
```

RESULT 5
BM021081/c
LOCUS BM021081 579 bp mRNA linear EST 30-OCT-2001
DEFINITION ie75f03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA 3' similar to TR:Q9Y231 Q9Y231 ALPHA-3-FUCOSYLTRANSFERASE. ;, mRNA sequence.
ACCESSION BM021081
VERSION BM021081.1 GI:16535437
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 579)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
High quality sequence stop: 436.

FEATURES Location/Qualifiers
source 1. .579
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"

/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 203 a 103 c 93 g 180 t
ORIGIN

Query Match 33.6%; Score 307.4; DB 10; Length 579;
Best Local Similarity 99.7%; Pred. No. 2.6e-74;
Matches 308; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    607 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 TACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTAT 520

Qy    667 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GAGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAG 460

Qy    727 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CTAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAAC 400

Qy    787 TGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCT 846
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    399 TGGAGGAAGGATTTCACTGTAAATCTTCCACCATTTTGGGAATCACATGCATGTTTGGCT 340

Qy    847 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTTGGTAATTTAGAGAAATGGTTT 280

Qy    907 TGGAATTAA 915
          |||||||||
Db    279 TGGAATTAA 271
```

RESULT 6
CNS059ST/c
LOCUS CNS059ST 1101 bp DNA linear GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 045P22 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL327638
VERSION AL327638.1 GI:8221227
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="045P22"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA045DH11A1-end : T3"
BASE COUNT 287 a 249 c 269 g 278 t 18 others
ORIGIN

Query Match 31.3%; Score 286; DB 12; Length 1101;
Best Local Similarity 59.1%; Pred. No. 2.7e-68;
Matches 537; Conservative 0; Mismatches 366; Indels 5; Gaps 3;

```
Qy     12 TTATTTTAATGAAACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCT 71
          | ||   ||    || | ||| ||| ||||| | ||||| || || || ||||| |||||
Db    941 TGATAAAAACATGACCATTATACTGCTGTGGTTCTGGCCGTTCGGACAAACCTTCGACCT 882

Qy     72 TACATCCTGCCAAGCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACT 131
          |     ||| | ||    ||| | | | || ||  |  | ||  ||||| |    ||
Db    881 GAACGTGTGCAACGCTTATTTCGGCTTTGAGGGCTGTTACATTACGGCGGACAGAAACCT 822

Qy    132 GTACAACAAATCCCATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAA 191
          |  ||||| ||  |||  ||| |  |||| ||| ||||||||    | || ||| | ||
Db    821 TTTTAACAAGTCGGATGGCGTTGTCTTCCACCACAGAGACATCGCCGGTGACCTGTCCAA 762

Qy    192 TTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACC 251
          | ||       |  | ||  || |||||||||||| ||| |  |  |||| || ||
Db    761 CATGCCGACAGTCGTGCGACCGTCCGTCCAGAAATGGATCTGGTTTNACATGGAGTCTCC 702

Qy    252 AACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCG 311
          | ||| |  |     |  ||| || |||  | |||| |||||||| |||| || ||
Db    701 CTCACACTCCTCTAGGCACTCTGGAATGGAGAGCCTGTTCAACCTGACGCTGAACTATCG 642

Qy    312 CCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGT 371
          |  ||||| || ||        || |  || |     ||  ||| || |    |
Db    641 TCTAGATTCTGACATTTCCACTGCTCTTTATTTATCCGTTGTTGCCACTGATGCTGAGGA 582

Qy    372 G---TTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCC 428
          |   ||||     || |||||| | ||| |||| || ||| |||| || ||||||||||
Db    581 GGACTTTGTCCCACCTAGCAAAAACAAACTGGTTTGTTGGATTGTCAGCAACTGGAACCA 522

Qy    429 TGAGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTA 488
          || |||||  |||| || |  ||||| ||  |   |||      |||| |||| || ||
Db    521 GGAACATGCACGAGTTAAATTCTACAACGATTTGTACAAGCATGTTGAGGTCCACACGTA 462

Qy    489 CGGGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTG 548
          || | ||||||  ||  || |  ||   ||| || |     | || |||||  |    ||
Db    461 TGGACGAGCATTCGGGCAACACTTCCCAGATGAAGAGACCTTACCAACCATCGCCAGCTG 402

Qy    549 TAAATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATA 608
          || || || || || ||||| || || |||||||| || ||| | || ||||| || ||
Db    401 CAAGTTCTACCTGTCATTTGAGAACTCCATCCACAAAGACTACTTTACTGAAAAACTGTA 342

Qy    609 CAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGA 668
          |||  |  ||   |  ||  | || || || |||||||| ||| || |||| ||||| ||
Db    341 CAACCCACTTGCTGTGGGGACAGTGCCAGTGGTTCTGGGCCCACCTCGGGAGAACTACGA 282

Qy    669 GAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCT 728
          |||    || |  | |||| | || || || |||||||| | |  |||  |||  |||||
```

```
Db    281 GAACATCATCCACGGAGATGCCTTTATCCACGTGGAAGACTTTGCCTCAGCCAAGGAGCT 222

Qy    729 AGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 788

Db    221 TGCTGACTATTTACTGCTCCTGGACAAAAATGAGGAAATGTACCTTAAATATTTTCANGT 162

Qy    789 GAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTTG 848

Db    161 GGCGAAGACACTTTAAAGCAAGGGCAGTGGCATTTGCCCGAGGC-CACCTGTCAGGCTTG 103

Qy    849 CGATCATGTGAAAAGGCATCAAGAATATA-AGTCTGTTGGTAATTTAGAGAAATGGTTTT 907

Db    102 TGATTACGTCAGGAGGCNTNANCAATACAGAGTATTCAGTGACCTTANACAAGTGGTACT 43

Qy    908 GGAATTAA 915

Db     42 GGGGTTGA 35
```

```
RESULT 7
BM082484/c
LOCUS       BM082484                570 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  fu25e08.x1 Campbell zebrafish ovary Danio rerio cDNA clone 5306223
            3' similar to TR:O88819 O88819 ALPHA1,3-FUCOSYLTRANSFERASE IX. ;,
            mRNA sequence.
ACCESSION   BM082484
VERSION     BM082484.1  GI:16929414
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fu25e08.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library constructed by Library constructed by Invitrogen and
            donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5306223"
                     /clone_lib="Campbell zebrafish ovary"
                     /sex="female"
                     /dev_stage="4-5 months, 1 year and 2 years"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Whole ovaries collected from zebrafish
                     aged 4-5 months, 1 year and 2 years.  Oligo-dT primed,
                     directionally cloned.  Average insert size 2 kb.  Library
                     constructed by Invitrogen and donated by R. Campbell
                     (Marine Biology Laboratory, Woods Hole, MA)."
BASE COUNT      155 a    106 c    129 g    180 t
ORIGIN
```

```
  Query Match           20.1%;  Score 184.2;  DB 10;  Length 570;
  Best Local Similarity 59.2%;  Pred. No. 3.6e-40;
  Matches  337;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    1;

Qy    217 TTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGC 276

Db    570 TTCCAGAAGTGGATTTGGTTACATTTAGAATCGCCGACCAACACTAAAATAATCCCTGGT 511

Qy    277 ATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCT 336

Db    510 CTGGAAAACCTGTTCAACCACACTCTCAGCTACAGACAAGATGCTGATATTCCAGTACGG 451

Qy    337 TATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAA 396

Db    450 ATGAGATTAAAAACCAGGAAGAAACCAGCCGATGATTTTGTCATTCCCAAAAAGGACAAA 391

Qy    397 TTGGTGTGCTGGGTTGTGAGTAACTGGAACCC---------TGAGCATGCCAGAGTCAAG 447

Db    390 CTAGTTTGTTGGATTGTGAGTAACAATTCCCCATTAACAGGTGTCAGCGCAAGGAACAGG 331

Qy    448 TATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAA 507

Db    330 ATTTATCGGGAATTAAGCCGATACATCCATGTACACCTTTTCGGGAAGGCTTATTCAACA 271

Qy    508 TATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTT 567

Db    270 TATCTAGACTATAAAGACTACTATCCTACCCTCGCTAGCTGCAAATTTTACCTCTCCTTC 211

Qy    568 GAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGC 627

Db    210 GAGAACTCCATCCACAAGGACTACATCACTGAGAAGATCAATGGTCCTCTTGCTGCAGGA 151

Qy    628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687

Db    150 ACTGTTCCTGTGGTTTTGGGTCCTCCAAGGAAAAACTATGAAAACTTCATTCCTGGAGAC 91

Qy    688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 747

Db     90 GCCTTTATTCATGTGAATGACTTCCTAAGTGTCAGGTCTCTAGCACGATATTTGTATATG 31

Qy    748 GTCGACAAAAACAATAAGTTATACCTTAG 776

Db     30 CTAAATAAAAATGACATTGCATATCACAG 2
```

```
RESULT 8
CNS0579Z/c
LOCUS       CNS0579Z                1100 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            007H05 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL324368
VERSION     AL324368.1  GI:9557250
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
```

```
          freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007H05"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA007CD03C1-end : T7"
BASE COUNT      314 a    228 c    252 g    304 t      2 others
ORIGIN

  Query Match           19.3%;  Score 177;  DB 12;  Length 1100;
  Best Local Similarity   55.4%;  Pred. No. 4.3e-38;
  Matches  428;  Conservative    0;  Mismatches  332;  Indels   13;  Gaps    4;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| | ||| | ||||| | ||| |||| |||| |||| || ||  |||
Db     773 ATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTTGCATACGACTCTTGTAAAGTTTAC 714

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           | ||| ||  | || ||  | |||||   || || || || ||| ||||| | || ||
Db     713 TACAATATTGACGGCTGTGAGCTCACAGATGATCGATCCCTTTACGACAAAGCGCAGGCT 654

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           |||||| ||   |||   |||||    ||| |||||  ||| || ||   | |  |  ||
Db     653 GTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGGAAACTTGCCAGTGGAGCCGCGG 594

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           |||   || ||||  |||||||||| |  |  ||||||| || |   ||||   |   |
Db     593 CCATATTTTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGATA 534

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           |||  | ||     | || ||| |||||||||  ||| |    ||||| || ||   |
Db     533 CCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGTA 474

Qy     331 GTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 387
           |    ||   |    | | |   |     | |   |        | |   ||||
Db     473 GCCAGGTACCCTCTTACCATCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGAG 414

Qy     388 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGC------ATGCCAGA 441
           ||| | ||| | |  || ||| | || || ||    |   |            | | |||
Db     413 AAAAACAAACTTGCCTGTTGGATCGTTAGCAATGCTACATCCACTGGGACAGGTACAAGA 354

Qy     442 GTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTT 501
           |   | | ||| ||||| |||  ||||  ||| | ||   |   |  ||   | |
Db     353 GCGCAATTTTATAATGAACTATCCAAACATATTAACATTAGTGTTTTTGGTGTGGTGTAC 294

Qy     502 GGAGAATATGTCAATGATAAAAATTTGATT---CCTACCATATCTGCTTGTAAATTTTAT 558
             ||          || |   ||  |  |   || ||| |  ||   ||||| |||||
Db     293 ACAGGGGTCAGGTTGGAGATTGATCAGTATTACCCGACCGTTGCTAGCTGTAAGTTTTAC 234

Qy     559 CTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCTTTT 618
           | || ||||| || ||| | ||||| ||||||||||| || ||| |  ||    |  ||
Db     233 TTATCATTTGAGAACTCATTACACAAAGATTACATCACAGAGAAGGTCAACGNACCCCTT 174

Qy     619 CTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATT 678
           || ||  | || ||||| ||||||||||  | || |  ||||| || |  |   |
Db     173 GCAGCAGGAACCGTTCCTGTAGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTTC 114

Qy     679 CCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTAT 738
           ||  |||| || || ||||| || || |||| |     ||| |  |  || ||| || ||
Db     113 CCTTCAGACTCCTTTATTCACGTAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCAT 54
```

```
Qy     739 CTGAAGGAAGTCGACAA-AAACAATAAGTTATACCTTAGTTACTTTAACTGGA 790
           |||       | |||||   |  |  ||  |||| |  | |||||| | ||||
Db      53 CTGCTCTCCTTGGACAAGGNATGAGGAGGCATACATGCGCTACTTTGAGTGGA 1


RESULT 9
CNS0429P/c
LOCUS       CNS0429P                1016 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            075P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271222
VERSION     AL271222.1  GI:7993186
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1016)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1016)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1016
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="075P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG075DH01SP1-end :
                     PUC-Ori"
BASE COUNT      246 a    216 c    253 g    298 t      3 others
ORIGIN

  Query Match           18.6%;  Score 170.2;  DB 12;  Length 1016;
  Best Local Similarity   57.0%;  Pred. No. 3.2e-36;
  Matches  394;  Conservative    0;  Mismatches  283;  Indels   14;  Gaps    4;

Qy      32 TTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATGT 91
           ||||| ||||| | ||||| | ||  |  |||||  | |     |||    |    |
Db     715 TTCTGCTGTGGTTCTGGCCCATCGGCACGGTGTTTGATTTCAAGGATTGCTCCACCCACT 656

Qy      92 TCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAG 151
           |||| ||  |  | || ||||| ||     ||| | || ||||||| |||| |  | | ||
Db     655 TCAAGATTGACAGCTG-CATCTGACGGACGACAGATCTCTGTACAGCAAAGCAAAGGGAG 597

Qy     152 TTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGC 211
           |  |  || | |||  |||||  |  ||| |       |  || ||   |    | | |
Db     596 TCATTTTCTACCACAAAGACACAACATGGCACACAGACAGCTTGCCCACGGGTCCACGCC 537

Qy     212 CACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGA 271
           | |  |||||||  ||| ||||| | ||| |||||||| |||||  ||||   |   ||
Db     536 CCCATTTCCAGAGGTGGGTTTGGTTTAATGTGGAATCCCCAACAAACACATACAGGAAAC 477
```

```
Qy     272 GTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAG 331
Db     476 CAGGTCTGGAAAACCTGTTCAACCTGACGCTGAGCTACAGGAGAGACGGCGGCATTGTGG 417

Qy     332 TGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTG---TTTGAAGTGCCAAGCA 388
Db     416 CGAGGAACGAGATGTCTATCCGAGACCCAGGTGACACCGAGGACTTTGTTCTGCCCAAGA 357

Qy     389 AAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATG---------CCA 439
Db     356 AAGACAAGACGGTGTGTTGGATCGTCAGCAACAATGACCCGAGAACAGGAACAGGCACCA 297

Qy     440 GAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCAT 499
Db     296 GAGACCAATACTACCGTGACCTGTCCAAACACATCAAGGTGGACGTTTGGGGCAAAGCCT 237

Qy     500 TTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATC 559
Db     236 TTGGGAATTTCCTGAGTTTTGAGGAATACTATCCAACAATTGCCAGCTGTAAATTCTATC 177

Qy     560 TTTCCTTTGAAAATTCAATCCACAAGGATTACATCACCGAAAAGCTATACAATGCTTTTC 619
Db     176 TGTCTTTCGAGAACTCCATCCACAAAGACTACATGACGGAGAAGATCAACGGGCCCTTTG 117

Qy     620 TGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATAT-T 678
Db     116 TGGCTGGGTCAGTTCCTATAGTTTTGGGACCTCCAAAAAAAAAAAAAAAAAAAATTCTAT 57

Qy     679 CCAGCAGATTCATTCATTCATGTGGAAGATT 709
Db      56 CCTTCCGATTCGTTTATTCACGTTAACGACT 26
```

RESULT 10
CNS01V8V

```
LOCUS       CNS01V8V                 844 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            198K09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL168808
VERSION     AL168808.1  GI:7806865
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 844)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="198K09"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG198AF05SP1-end :
                     PUC-Ori"
BASE COUNT      199 a    190 c    205 g    245 t      5 others
ORIGIN
```

```
  Query Match             17.6%;  Score 161.4;  DB 12;  Length 844;
  Best Local Similarity   56.2%;  Pred. No. 8.3e-34;
  Matches  319;  Conservative    2;  Mismatches  244;  Indels    3;  Gaps    1;

Qy      25 ACTACTATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAA 84
Db     280 ACCACGGTGCTGCTCTGGTTCTGGCCATTTGGGATTACCTTTGACCCGAACGTGTGCAGC 339

Qy      85 GCAATGTTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCC 144
Db     340 GCTTACTATGGTATAGACGGCTGTTTCATTACAGCGGACAGGAACTATTTCAGCGAAGGA 399

Qy     145 CATGCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAA 204
Db     400 ATGCGGNTCGTTATTCATCATCGAGAMATTTCTCCCGACCTGTCCAATCTGCCGCAGCAG 459

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
Db     460 GAGCGGCCGTTGTACCARATGTGGGTCTGGTTCAACATGGAGAATCCTTCGTACTCCGAC 519

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
Db     520 AGGTTTCTTGGAATGGACGGCTTGTTCAACCTGACTCTGAATTATCGTCTGGATGCTGAT 579

Qy     325 ATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCA 384
Db     580 ATTTCCATGCCTCATGTGTACCTGGTTCCTGAAAAAAGTGAGGAAGACTTTGTCCCCCCC 639

Qy     385 AGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTC 444
Db     640 AGCAAAGATAAGTTGGTCTGCTGGATAGTCAGCAACTGGAAGGAGCATTATATACGGGTG 699

Qy     445 AAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGA 504
Db     700 AAATACTTCAAGGAGCTGTCCAAGCACATAAAGATCACCACTTATGGACGTGGCTTTGGG 759

Qy     505 GAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCC 564
Db     760 GGCTA---CAACTGGGGAAGTGCCTCGTCCATCATGGCCAGCTGTAAGTTCTACCTGTCT 816

Qy     565 TTTGAAAATTCAATCCACAAGGATTACA 592
Db     817 TTTGAGAACTCCATCTACATTGACTACA 844
```

RESULT 11
CNS02Q0S/c

```
LOCUS       CNS02Q0S                 887 bp    DNA     linear   GSS 14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            156L20 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL208693
VERSION     AL208693.1  GI:7867512
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
```

```
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 887)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 887)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .887
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="156L20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG156DF10LP1~end : T7"
BASE COUNT      232 a    188 c    206 g    253 t      8 others
ORIGIN

  Query Match             17.2%;  Score 157.4;  DB 12;  Length 887;
  Best Local Similarity   54.2%;  Pred. No. 1.1e-32;
  Matches  380;  Conservative    5;  Mismatches  301;  Indels   15;  Gaps    3;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| | ||| | ||||| | ||| |||| || |         || ||  |||
Db     703 ATCCTGTTATGGTTCTGGCCGTATGGACAGAAGTTGGCATACGACTCTTGTAAAGTTGAC 644

Qy      91 TTCAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCA 150
           : ||| ||  |  | ||  | ||:|||   || || || ||  || ||||| | || ||
Db     643 KACAATATTGACAGCTGTGAGCTSACAGATGATCGATCCCTTGACGACAAAGCGCAGGCT 584

Qy     151 GTTCTGATCCATCACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGG 210
           |||||| ||   |||   |||||    ||| ||||| |||| || ||   | |  |  ||
Db     583 GTTCTGTTCTTCCACAAGGACATACAGTGGAATCTGGCAAACTTGCCAGGGGAGCCGCGG 524

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           |||   |: ||||  |||||||||| |  |  ||||||| || |   ||||   |   |
Db     523 CCATATTKTCAGAGGTGGATTTGGTTCTACCTGGAATCTCCCAGGAACACAATCAGGATA 464

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
           |||  | ||     | || ||| |||||||||  ||| |    ||||| || ||
Db     463 CCAGGCCTGGAAACAGTTTTCAACATGACTCTGAACTACAGGAAAGATTCCGACATTGTG 404

Qy     331 GTGCCTTA---TGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGC 387
           |    ||   |    | | |  |    | |   |       | |   ||||
Db     403 GCCAGGTACCCTCTTACCATCAGAGAGGAAAGACTGACGGAGAAAATAGKTCTGCCGGAG 344

Qy     388 AAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCA-- 445
           ||| | ||| | |||||||||| | || || ||           |:  || ||  | ||
Db     343 AAAAACAAAATTGTGTGCTGGATCGTCAGCAAAACTGCAAAAAASACTGGCACGGGCACA 284

Qy     446 -------AGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCA 498
               | |  |    ||| ||| |||   ||| |  |  ||   |  ||    ||
Db     283 AGGGAACAATTCTTTGGTGAACTATCCAAGCATATTCATGTTGATGTTTTTGGTGTGGCG 224

Qy     499 TTT---GGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           ||||   |  |  |  | |  | |     || |||||  ||   ||||| |||
Db     223 GACACAGGAGTGAAGTTGGAGATTGATCAGTATTACCCGACCATTGCTAGCTGTAAGTTT 164

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTATACAATGCT 615
           || || |||||||| || |||   ||||| |||:||||||| || ||| |  ||    |
Db     163 TACCTGTCCTTTGAGAACTCATAACACAAAGATKACATCACAGAGAAGGTCAACGGACCC 104

Qy     616 TTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTAT 675
            ||   |  ||  | || ||||| ||||||||||  | || |  ||||| || |  |
```

```
Db     103 CTTGCAGTAGGAACCGTTCCTGTGGTTCTGGGACCCCCAAGAGCCAACTACGAAATGTTC 44

Qy     676 ATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTC 716
             | ||  |||| || || ||||| || || |||| |   ||
Db      43 TTCCCTTCAGACTCCTTAATTCACGTAGATGATTTTCTTTC 3


RESULT 12
BJ036855
LOCUS       BJ036855                651 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  BJ036855 NIBB Mochii normalized Xenopus neurula library Xenopus
            laevis cDNA clone XL037k06 5', mRNA sequence.
ACCESSION   BJ036855
VERSION     BJ036855.1  GI:17398869
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL037k06"
                     /clone_lib="NIBB Mochii normalized Xenopus neurula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 15"
BASE COUNT      191 a    157 c    119 g    181 t      3 others
ORIGIN

  Query Match             17.0%;  Score 155.4;  DB 10;  Length 651;
  Best Local Similarity   56.6%;  Pred. No. 3.5e-32;
  Matches  329;  Conservative    0;  Mismatches  246;  Indels    6;  Gaps    2;

Qy     211 CCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAG 270
           || ||    ||  | ||| ||||| | |  ||||| || ||||  |||| |||| | |
Db      74 CCCCCAAAACAATATTGGGTTTGGTTTACCTTGGAGTCCCCAAGCCACAATCCCAACACA 133

Qy     271 AGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAA 330
               || ||| |  |  | || |||||  || |||||||   ||| |||||||||
Db     134 GCCTTTATGGAGAAACTCATCAATCTGACCATGTCTTACCGTGCTGACTCAGATATTTTT 193

Qy     331 GTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAA 390
           ||||||||||  |                  ||      ||||   |||
Db     194 ACTCCTTATGGCTGGCTAGAAAGGCATGATGGAAAAGAAAACTTCACTATGCCCGCCAAG 253

Qy     391 GAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTAT 450
           ||| | ||   |||||| || ||||||||||||||  |       ||| |||| |||
Db     254 ACAAAACTAGTAGCCTGGGTAGTCAGTAACTGGAACCCAAACTCCAGAAGAATCAAATAT 313

Qy     451 TACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATAT 510
           ||  | ||| |||    |  | |||   |  ||| ||| || | |      |  |     |
Db     314 TATGAGGAGTTAAAGCCACACCTTGCTGTTGATATCTATGGCCTACAGCATTTACCGCTT 373

Qy     511 GTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAA 570
           |     ||| |    |||   |  | ||    | |||||||| |  | ||||||
Db     374 CCTAGACCAAAACA---CATTGAAATTCTTTCCCAGTACAAATTTTATTTAGCTTTTGAA 430

Qy     571 AATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGC 627
```

```
Db     431 AATTCAATTCACGAAGACTACATCACTGAGAAACTCTGGCATAATGCCCTCTCATCGTGG 490

Qy     628 TCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGAT 687

Db     491 GCTCTACCTGTTGTTTTGGGGCCCTCGCGTTCTAACTATGAGCGTTTTTTACCCCCCGAG 550

Qy     688 TCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAA 747

Db     551 TCATTCATCCACGTAGATGACTTTCCAACTGCTAAGGGACTTGCTGATTATCTAAACGAG 610

Qy     748 GTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTG 788

Db     611 CTTGATAAAGATGATAATAAATACATGCAACATTTCAACTG 651


RESULT 13
BJ050300/c
LOCUS       BJ050300                666 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  BJ050300 NIBB Mochii normalized Xenopus neurula library Xenopus
            laevis cDNA clone XL023o15 3', mRNA sequence.
ACCESSION   BJ050300
VERSION     BJ050300.1  GI:17392988
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL023o15"
                     /clone_lib="NIBB Mochii normalized Xenopus neurula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 15"
BASE COUNT      178 a    116 c    157 g    215 t
ORIGIN

  Query Match             16.6%;  Score 151.6;  DB 10;  Length 666;
  Best Local Similarity   54.5%;  Pred. No. 4e-31;
  Matches  348;  Conservative    0;  Mismatches  284;  Indels    6;  Gaps    2;

Qy     251 CAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACC 310

Db     666 CAAGCCACAATCCCAACACAGCGTTTATGGAGAAACTCATCAATCTGACCATGTCTTACC 607

Qy     311 GCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGACGGTAAGCACAAATCCCTTCG 370

Db     606 GTGCTGACTCAGATATTTTTACTCCTTATGGCTGGCTAGAAAGGCATGATGGAAAAGAAA 547

Qy     371 TGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTG 430

Db     546 ACTTCACTATGCCCGCCAAGACAAAACTAGTAGCCTGGGTAGTCAGTAACTGGAACCCAA 487

Qy     431 AGCATGCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACG 490

Db     486 ACTCCAGAAGAATCAAATATTATGAGGAGTTAAAGCCACACCTTGCTGTTGATATCTATG 427

Qy     491 GGCAAGCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTA 550

Db     426 GCCTACAGCATTTACCGCTTCCTAGACCAAAACA---CATTGAAATTCTTTCCCAGTACA 370

Qy     551 AATTTTATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT--- 607

Db     369 AATTTTATTTAGCTTTTGAAAATTCAATTCACGAAGACTACATCACTGAGAAACTCTGGC 310

Qy     608 ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATG 667

Db     309 ATAATGCCCTCTCATCGTGGGCTCTACCTGTTGTTTTGGGGCCCTCGCGTTCTAACTATG 250

Qy     668 AGAATTATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGC 727

Db     249 AGCGTTTTTTACCCCCCGAGTCATTCATCCACGTAGATGACTTTCCAACTGCTAAGGGAC 190

Qy     728 TAGCAAAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACT 787

Db     189 TTGCTGATTATCTAAACGAGCTTGATAAAGATGATAATAAATACATGCAACATTTCAACT 130

Qy     788 GGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTGGGAATCACATGCATGTTTGGCTT 847

Db     129 GGAGGAAACATCTAAAACCAGTAGCACCTGAGAGTTGGAACCCCCATTACTGCAAAGTAT 70

Qy     848 GCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGTT 885

Db      69 GCAAAGCTATAAAAGACGCACCGTCTTATAGGACTTTT 32


RESULT 14
CNS034XF/c
LOCUS       CNS034XF                711 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            212J18 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL228012
VERSION     AL228012.1  GI:7886987
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 711)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 711)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="212J18"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG212DE09SP1-end :
```

```
                    PUC-Ori"
BASE COUNT      185 a    153 c    168 g    203 t      2 others
ORIGIN

  Query Match             16.5%;  Score 151.2;  DB 12;  Length 711;
  Best Local Similarity   55.0%;  Pred. No. 5.3e-31;
  Matches  360;  Conservative    1;  Mismatches  284;  Indels    9;  Gaps    3;

Qy    170 ACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGAAATGGA 229
          ||||    ||| ||||| |||| || ||   | |  |  |||||   || |||| ||||
Db    705 ACATACAGTGGAATCTGGCAAACTTTCCAGTGGAGCCGCGGCCATATTTTCAGAGATGGG 646

Qy    230 TTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGCACTTGT 289
          ||||| |  || | ||| | ||||   ||||   |   |    |||  | ||      | |
Db    645 TTTGGTTCTATCTAGAACCTCCAATGAACACATACAGGATACCTGGTCTAGAAACACTCT 586

Qy    290 TTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCTTCTTGA 349
          | ||| ||||||||   ||| |    || || || ||    |     ||     |    |
Db    585 TCAACATGACTCTGCACTACAGGAAAGACTCCGACATTGTGGCCAGGTACCCTCTTACCA 526

Qy    350 CGGTAAGCACAAATCCCTTCGTGTTTGAAG---TGCCAAGCAAAGAGAAATTGGTGTGCT 406
               |     ||  |     | |     ||   ||||    ||| | |||   ||||| |
Db    525 TCAGAGAGGAAATACTGACGGAGAAAATAGTTCTGCCGGAGAAAAACAAAACTGTGTGTT 466

Qy    407 GGGTTGTGAG---TAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
          || | || ||   |||  | |          ||| |:|    | | |||   ||| |||
Db    465 GGATCGTCAGCAATAATCGTACATTCACTGGTGCAARACAGCAATTTTATCGTGAACTAT 406

Qy    464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCAT---TTGGAGAATATGTCAATGATA 520
            | |    ||| | ||  ||   || ||  | || |     || |      |  |   |
Db    405 CCCAGCATATTTACATAAATGTTTATGGTTACGCCTACACAGGGGTCAGGTTGGAGATTG 346

Qy    521 AAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCC 580
          |  | |   | |  ||| |  ||   ||||| ||||| || |||||||| || ||||||
Db    345 ATCAATATTTCCAGACCGTTGCTAGCTGTAAGTTTTACCTGTCCTTTGAGAACTCAATCT 286

Qy    581 ACAAGGATTACATCACGGAAAAGCTATACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
          |||| |||||||||||||| ||| |  ||    |  ||   |  ||  | || ||||| |
Db    285 ACAAAGATTACATCACGGAGAAGGTCAACGGACCCCTTGCAGTAGGAACCGTTCCTGTAG 226

Qy    641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
          ||||||||||  | || |  ||||| || |  |   | ||  |||| || || ||| |
Db    225 TTCTGGGACCCCCAAGAGCCAACTACGAAATGTTCTTCCCTTCAGACTCCTTTATTAACA 166

Qy    701 TGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTCGACAAAAACA 760
          | || |||| |     ||| |  |  || ||| || |||||       | |||||  |
Db    165 TAGATGATTTTGCAGATCCAAAGGCACTGGCAGAGCATCTGCTCTCCTTGGACAAGGATG 106

Qy    761 ATAAGTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTC 814
          |  ||  |||| |  | |||||| | ||||||||| | ||||  || | || ||
Db    105 AGGAGGCATACATGCGCTACTTTGAGTGGAGGAAGTACTTCAGCGTCATTCCTC 52


RESULT 15
B41073/c
LOCUS       B41073                   501 bp    DNA     linear   GSS 18-OCT-1997
DEFINITION  HS-1052-B2-G01-MF.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 774 Col=2 Row=N, DNA sequence.
ACCESSION   B41073
VERSION     B41073.1  GI:2545325
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 774  row: N  column: 2
            Class: BAC ends
            High quality sequence stop: 501.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=CT 774 Col=2 Row=N"
                     /clone_lib="CIT Human Genomic Sperm Library C"
                     /sex="M"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      179 a     95 c     83 g    143 t      1 others
ORIGIN

  Query Match             16.5%;  Score 151;  DB 12;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-31;
  Matches  151;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    765 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 GTTATACCTTAGTTACTTTAACTGGAGGAAGGATTTCACTGTAAATCTTCCACGATTTTG 442

Qy    825 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 GGAATCACATGCATGTTTGGCTTGCGATCATGTGAAAAGGCATCAAGAATATAAGTCTGT 382

Qy    885 TGGTAATTTAGAGAAATGGTTTTGGAATTAA 915
          ||||||||||||||||||||||||||||||
Db    381 TGGTAATTTAGAGAAATGGTTTTGGAATTAA 351


Search completed: October  6, 2002, 07:16:27
Job time : 773.105 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 22:42:00 ; Search time 18.6694 Seconds
(without alignments)
12038.636 Million cell updates/sec

Title: US-09-744-748-5_COPY_460_1374
Perfect score: 915
Sequence: 1 accaaaactgattattttaa.....agaaatggttttggaattaa 915

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	13.9	1814	2	US-08-483-151-1	Sequence 1, Appli
2	127	13.9	1814	5	PCT-US96-06427-1	Sequence 1, Appli
3	113.4	12.4	1654	1	US-07-914-281-13	Sequence 13, Appl
4	113.4	12.4	1654	1	US-08-393-246-13	Sequence 13, Appl
5	113.4	12.4	1654	1	US-08-525-058A-13	Sequence 13, Appl
6	113.4	12.4	1654	2	US-08-696-731-13	Sequence 13, Appl
7	113.4	12.4	1654	4	US-09-042-531-13	Sequence 13, Appl
8	112.2	12.3	1086	1	US-07-914-281-12	Sequence 12, Appl
9	112.2	12.3	1086	1	US-08-393-246-12	Sequence 12, Appl
10	112.2	12.3	1086	1	US-08-525-058A-12	Sequence 12, Appl
11	112.2	12.3	1086	2	US-08-696-731-12	Sequence 12, Appl
12	112.2	12.3	1086	4	US-09-042-531-12	Sequence 12, Appl
13	112.2	12.3	2042	3	US-09-063-237-2	Sequence 2, Appli
14	112.2	12.3	2043	1	US-07-914-281-1	Sequence 1, Appli
15	112.2	12.3	2043	1	US-08-393-246-1	Sequence 1, Appli
16	112.2	12.3	2043	1	US-08-273-411-4	Sequence 4, Appli
17	112.2	12.3	2043	1	US-08-525-058A-1	Sequence 1, Appli
18	112.2	12.3	2043	2	US-08-696-731-1	Sequence 1, Appli
19	112.2	12.3	2043	4	US-09-042-531-1	Sequence 1, Appli
20	112.2	12.3	2043	5	PCT-US91-00899-1	Sequence 1, Appli
21	111.8	12.2	1316	1	US-07-914-281-10	Sequence 10, Appl
22	111.8	12.2	1316	1	US-08-393-246-10	Sequence 10, Appl
23	111.8	12.2	1316	1	US-08-525-058A-10	Sequence 10, Appl
24	111.8	12.2	1316	2	US-08-696-731-10	Sequence 10, Appl
25	111.8	12.2	1316	4	US-09-042-531-10	Sequence 10, Appl
26	100.8	11.0	1068	5	PCT-US91-00899-13	Sequence 13, Appl
27	100.8	11.0	1215	5	PCT-US91-00899-5	Sequence 5, Appli
28	100.8	11.0	1488	1	US-07-914-281-9	Sequence 9, Appli
29	100.8	11.0	1488	1	US-08-393-246-9	Sequence 9, Appli
30	100.8	11.0	1488	1	US-08-525-058A-9	Sequence 9, Appli
31	100.8	11.0	1488	2	US-08-696-731-9	Sequence 9, Appli
32	100.8	11.0	1488	4	US-09-042-531-9	Sequence 9, Appli
33	100.8	11.0	2175	4	US-08-482-073-9	Sequence 9, Appli
34	100.8	11.0	2861	4	US-08-482-073-10	Sequence 10, Appl
35	100.8	11.0	3647	1	US-07-914-281-7	Sequence 7, Appli
36	100.8	11.0	3647	1	US-08-393-246-7	Sequence 7, Appli
37	100.8	11.0	3647	1	US-08-525-058A-7	Sequence 7, Appli
38	100.8	11.0	3647	2	US-08-696-731-7	Sequence 7, Appli
39	100.8	11.0	3647	4	US-09-042-531-7	Sequence 7, Appli
40	100.8	11.0	3647	5	PCT-US91-00899-4	Sequence 4, Appli
41	97.6	10.7	2134	2	US-08-483-151-3	Sequence 3, Appli
42	97.6	10.7	2134	5	PCT-US96-06427-3	Sequence 3, Appli
c 43	37.2	4.1	1728	1	US-08-403-866-12	Sequence 12, Appl
c 44	37.2	4.1	2231	1	US-08-403-866-14	Sequence 14, Appl
c 45	37.2	4.1	12720	1	US-08-403-866-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-483-151-1
; Sequence 1, Application US/08483151
; Patent No. 5858752
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Holgersson, Jan
; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,151
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/278001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1814 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-483-151-1

Query Match 13.9%; Score 127; DB 2; Length 1814;
Best Local Similarity 52.3%; Pred. No. 9.7e-29;
Matches 330; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

```
Qy    104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
          | |||| ||| |   |  ||||    |||     ||   |  |||| ||  |  |||| |
Db    548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607
```

```
Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           |||| ||  |          |    |     |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
             ||| | |||       ||||||| || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           | | || |||  |   |||| || || ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
           ||||  |     | |    ||   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           ||||||  | || || |  |     |||   | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
           |       |  |  |  ||   | ||| |  ||     |||       |   | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
           | || | || ||  |  |     |      || || ||  |||||| || ||| |  ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           ||| |||||||| || ||| | |     ||||||  |   ||||||  |||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||  || |||   | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           |||| || |  | |||| || |||| || ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 2
PCT-US96-06427-1
; Sequence 1, Application PC/TUS9606427
;  GENERAL INFORMATION:
;    APPLICANT:  The General Hospital Corporation
;    TITLE OF INVENTION:  FUCOSYLTRANSFERASE GENES AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/483,151
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/278WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1814 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-06427-1

  Query Match             13.9%;  Score 127;  DB 5;  Length 1814;
  Best Local Similarity   52.3%;  Pred. No. 9.7e-29;
  Matches  330;  Conservative    0;  Mismatches  295;  Indels    6;  Gaps    2;

Qy     104 GATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATC 163
           | |||| ||| |   |  ||||    |||      ||    |  |||| ||  |  |||| |
Db     548 GCTGCCGTCTGAGTGCTAACCGGAGCCTGCTAGCCAGTGCTGATGCTGTGGTCTTCCACC 607

Qy     164 ACCGAGACATCAGTTGGGATCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAGA 223
           ||||  ||  |          |    |     |||| | | |    ||||||| |   |||
Db     608 ACCGTGAGCTGCAAACCCGGCAATCTCTCCTACCCCTGGACCAGAGGCCACACGGACAGC 667

Qy     224 AATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAGC 283
             ||| | |||       ||||||| || | | | || |          | || |   |
Db     668 CTTGGGTCTGGGCCTCCATGGAATCGCCCAGTAATACCCATGGTCTCCATCGCTTCCGGG 727

Qy     284 ACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGCT 343
           | | || |||  |   |||| || || ||||||||||||||||   || || || ||
Db     728 GCATCTTCAACTGGGTGCTGAGCTATCGGCGTGATTCAGATATCTTTGTACCCTACGGTC 787

Qy     344 TCTTGACGGTAAGCACAAATCCCTTCGTGTTTGAAGTGCCAAGCAAAGAGAAATTGGTGT 403
           ||||  |     | |    ||   |    | | ||   ||||   |   |||
Db     788 GCTTGGAGCCTCTCTCTGGGCC---CACATCCCCACTACCGGCCAAAAGCAGGATGGCTG 844

Qy     404 GCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTACAATGAGCTAA 463
           ||||||  | || || |  |     |||   | |  |||   |||    ||||
Db     845 CCTGGGTGATCAGCAATTTCCAGGAGCGGCAGCAGCGTGCAAAGCTGTACCGGCAGCTGG 904

Qy     464 GCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTCAATGATAAAA 523
           |       |  |  |  ||   | ||| |  ||     |||       |   | |||
Db     905 CCCCTCATCTGCAGGTGGATGTGTTCGGTCGCGCCAGCGGACGGCCCCTATGCGCTAATT 964

Qy     524 ATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAATTCAATCCACA 583
           | || | || ||  |  |     |      || || ||  |||||| || ||| |  ||
Db     965 GTCTGCTGCCCACTTTGGCCCGGTACCGCTTCTACCTGGCCTTTGAGAACTCACAGCATC 1024

Qy     584 AGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCTGTACCTGTTG 640
           ||| |||||||| || ||| | |     ||||||  |   ||||||  |||||| || |
Db    1025 GGGACTACATCACTGAGAAGTTCTGGCGCAATGCCCTGGCGGCTGGTGCTGTACCCGTGG 1084

Qy     641 TTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCATTCATTCATG 700
           ||||||||  || |||   | ||| |||  || | | ||| |||||  | ||  | || |
Db    1085 CGCTGGGACCTCCTCGGGCCACCTACGAGGCTTTTGTGCCACCAGATGCCTTTGTACACG 1144

Qy     701 TGGAAGATTATAACTCTCCCAGTGAGCTAGC 731
           |||| || |  | |||| || |||| || ||
Db    1145 TGGACGACTTCAGCTCTGCCCGTGAACTGGC 1175


RESULT 3
US-07-914-281-13
; Sequence 13, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-13

  Query Match             12.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||     |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
            |       ||      |  ||| |  ||||  | |||||      ||||      |  | |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | ||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||           | | | |       |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||      |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||           |  |  |  |   ||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
             |      |      |  ||   ||||    ||  | ||     |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||    || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |   |    ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| |||  | |||||  || |       ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           ||
Db    1095 GAG 1097
```

```
RESULT 4
US-08-393-246-13
; Sequence 13, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-13

  Query Match             12.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.2e-24;
```

```
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |     ||      |  ||| |  ||||  | |||||      ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||           | | | |      |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
             |  |  |  |||      |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           |||||  ||    || ||| |     ||        |  |  |    |  |||||    |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
             |      |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || ||||||||  || |||||  |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||     ||    |||||  ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | |||  | ||  | ||||| || ||||| || |   |     ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| ||||  | |||||  || |       ||||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           |||
Db    1095 GAG 1097


RESULT 5
US-08-525-058A-13
; Sequence 13, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-13

  Query Match             12.4%;  Score 113.4;  DB 1;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |     ||      |  ||| |  ||||  | |||||      ||||     |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||     |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||           | | | |      |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
             |  |  |  |||      |  |||||  |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           |||||  ||    || ||| |     ||        |  |  |    |  |||||    |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
             |      |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || ||||||||  || |||||  |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914
```

```
Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| |||| 
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || || 
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| |||| | |||||  || |       |||| || |||||| |||| || 
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
             ||
Db    1095 GAG 1097
```

RESULT 6
US-08-696-731-13
; Sequence 13, Application US/08696731
; Patent No. 5955347
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,731
; FILING DATE: 14-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/393,246
; FILING DATE:
; APPLICATION NUMBER: US 08/220,433
; FILING DATE: 30-MAR-1994
; APPLICATION NUMBER: US 07/914,281
; FILING DATE: 20-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-696-731-13

Query Match 12.4%; Score 113.4; DB 2; Length 1654;
Best Local Similarity 51.2%; Pred. No. 1.2e-24;
Matches 401; Conservative 0; Mismatches 361; Indels 21; Gaps 5;

```
Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| ||||| |||||| |||    |  || | |  ||  |   ||||  ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |      ||      |  ||| |  ||||  | |||||      ||||      |  | |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || ||||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | ||||||  | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||    ||| || ||||          | | | |      |    | |     | |
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||     |  ||||| |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||            |  |  |  |    |||||  |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           |    |   |    |  ||  ||||   ||  | ||   |  |    |||| ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || |||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| |||| 
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |    |    |||| || || || 
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| |||| | |||||  || |       |||| || |||||| |||| || 
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
             ||
Db    1095 GAG 1097
```

RESULT 7
US-09-042-531-13
; Sequence 13, Application US/09042531
; Patent No. 6268193
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor

```
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/042,531
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/393,246
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/220,433
;       FILING DATE:  30-MAR-1994
;       APPLICATION NUMBER:  US 07/914,281
;       FILING DATE:  20-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavalleye, Jean-Paul M. P.
;       REGISTRATION NUMBER:  31,451
;       REFERENCE/DOCKET NUMBER:  2363-060-55
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1654 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-042-531-13

  Query Match             12.4%;  Score 113.4;  DB 4;  Length 1654;
  Best Local Similarity   51.2%;  Pred. No. 1.2e-24;
  Matches  401;  Conservative    0;  Mismatches  361;  Indels   21;  Gaps    5;

Qy      31 ATTCTGGTGTGGGTGTGGCCATTTGGGCAGACCTTTGACCTTACATCCTGCCAAGCAATG 90
           || ||| |||||  |||||| |||    |  || | |  ||  |   |||| ||  |||
Db     318 ATCCTGCTGTGGACGTGGCCTTTTAACAAACCCATAGCTCTGCCCCGCTGCTCAGAGATG 377

Qy      91 TT---CAACATCCAAGGATGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCAT 147
             |      ||      |  ||| |  |||| | |||||     ||||      |  |  |
Db     378 GTGCCTGGCACGGCTGACTGCAACATCACTGCCGACCGCAAGGTGTATCCACAGGCAGAC 437

Qy     148 GCAGTTCTGATCCATCACCGAGACATCAGTTGGGATCTGA---CAAATTTACCTCAGCAA 204
           || ||  |  | || |||||||  |||  |   | |  |   |  |  | || |
Db     438 GCGGTCATCGTGCACCACCGAGAGGTCATGTACAACCCCAGTGCCCAGCTCCCACGCTCC 497

Qy     205 GCTAGGCCACCCTTCCAGAAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCC 264
           | ||||  |     |||  |||||| ||| | |   |||| || ||||  |||
Db     498 CCGAGGCGGCAGGGGCAGCGATGGATCTGGTTCAGCATGGAGTCCCCAAGCCACTGCTGG 557

Qy     265 CAAAAGAGTGGCATTGAGCACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGAT 324
           ||   ||  | ||| ||    |  || || || ||  || | |||||| | || || ||
Db     558 CAGCTGAAAGCCATGGACGGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGAC 617

Qy     325 ATCCAAGTGCCTTATGGCT---------TCTTGACGGTAAGCACAAATCCCTTCGTGTTT 375
           |||     ||| || ||||           | | | |      |    | |     ||
Db     618 ATCTTCACGCCCTACGGCTGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTC 677

Qy     376 GAAGTGCCAAGCAAAGAGAAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCAT 435
           |  |  |   |||      |  ||||| |||||  |||   ||||||   ||  |
Db     678 AACCTCTCGGCCAAGACCGAGCTGGTGGCCTGGGCAGTGTCCAACTGGGGGCCAAACTCC 737

Qy     436 GCCAGAGTCAAGTATTACAATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAA 495
           ||||| ||    || ||| |    ||           |  |  |  |    |||||   |
Db     738 GCCAGGGTGCGCTACTACCAGAGCCTGCAGGCCCATCTCAAGGTGGACGTGTACGG---A 794

Qy     496 GCATTTGGAGAATATGTCAATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTT 555
           |  .  |      |      |   ||   ||||    ||  | ||    |  || ||
Db     795 CGCTCCCACAAGCCCCTGCCCCAGGGAACCATGATGGAGACGCTGTCCCGGTACAAGTTC 854

Qy     556 TATCTTTCCTTTGAAAATTCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAAT 612
           |||||  |||| || || ||  | |||   || ||||||||| || ||||| |     ||
Db     855 TATCTGGCCTTCGAGAACTCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAAC 914

Qy     613 GCTTTTCTGGCTGGCTCTGTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAAT 672
           ||  |   |||  |  | || || || || ||||| ||    ||    ||||| ||||
Db     915 GCCCTGGAGGCCTGGGCCGTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGG 974

Qy     673 TATATTCCAGCAGATTCATTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCA 732
           |   | ||| | ||  | ||||| || ||||| || |   |     ||||  || || ||
Db     975 TTCCTGCCACCCGACGCCTTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCC 1034

Qy     733 AAGTATCTGAAGGAAGTCGACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGG 792
             ||| ||| ||||  | |||||  || |       |||| || ||||||  |||| ||
Db    1035 CGGTACCTGCAGGAGCTGGACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGG 1094

Qy     793 AAG 795
           |||
Db    1095 GAG 1097


RESULT 8
US-07-914-281-12
; Sequence 12, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-914-281-12

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;
```

```
Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |    |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |   | |     |          | |  | |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |  ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |   |    |||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |       ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 9
US-08-393-246-12
; Sequence 12, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-393-246-12

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |    |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
          |||||  ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
          |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |   | |     |          | |  | |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||           |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |  ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||    ||    ||||| |||| |   | ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |   |    |||  || || ||   ||| ||| ||||  |
```

```
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |       |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 10
US-08-525-058A-12
; Sequence 12, Application US/08525058A
; Patent No. 5770420
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,058A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-525-058A-12

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 1086;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
            | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
              ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |   |  |   |     |          | |  |  |  |   |||

Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||           |  |  |  |    |||||  |   |              |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||     |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || || ||||| ||    ||    |||||  ||||  |   |  ||| | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          ||||| || ||||| || |   |    ||||  || || ||   ||| ||| ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |       |||||  || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 11
US-08-696-731-12
; Sequence 12, Application US/08696731
; Patent No. 5955347
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,731
;      FILING DATE:  14-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
```

```
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-696-731-12

  Query Match            12.3%;  Score 112.2;  DB 2;  Length 1086;
  Best Local Similarity  51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
           |  ||  |    | |   |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
           |   ||             |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
           || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
           ||  | |||   || ||||||||| || ||||| |     || ||  |   |||  |  |
Db    748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
           || || || || ||||| ||    ||    |||||  ||||   |   | |||  | ||  |
Db    808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
           ||||| || ||||| || |   |    ||||  || || ||   ||| |||  ||||  |
Db    868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  || |      ||||| || ||||||  |||| || ||
Db    928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972


RESULT 12
US-09-042-531-12
; Sequence 12, Application US/09042531
; Patent No. 6268193
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-042-531-12

  Query Match            12.3%;  Score 112.2;  DB 4;  Length 1086;
  Best Local Similarity  51.6%;  Pred. No. 2.2e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||    |  |  |  | ||  |  | || |||
Db    271 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 330

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
           | || ||||  |   |   |  | ||    | || |       | |||||  |     |||
Db    331 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 390

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    391 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 450

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
             |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    451 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 510

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
           |  ||  |    | |   |     |           | |  |  |  |   |||
Db    511 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 570

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
           |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    571 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 630

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
           |   ||             |  |  |  |    |||||  |   |                |
Db    631 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 690
```

```
Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
Db     691 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 747

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
Db     748 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 807

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
Db     808 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 867

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
Db     868 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 927

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
Db     928 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 972
```

```
RESULT 13
US-09-063-237-2
; Sequence 2, Application US/09063237
; Patent No. 6124267
;  GENERAL INFORMATION:
;    APPLICANT:  McEver, Rodger P.
;    APPLICANT:  Cummings, Richard D.
;    TITLE OF INVENTION:  O-Glycan Inhibitors of Selectin Mediated
;    TITLE OF INVENTION:  Inflammation Derived from PSGL-1
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center, 1201 West Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
;      ZIP:  30306-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,237
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/649,802
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF110CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)873-8794
;      TELEFAX:  (404)873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2042 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-063-237-2

  Query Match             12.3%;  Score 112.2;  DB 3;  Length 2042;
  Best Local Similarity   51.6%;  Pred. No. 3e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy     106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
Db     343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy     166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
Db     403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy     223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
Db     463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy     283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
Db     523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy     343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
Db     583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy     394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
Db     643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy     454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
Db     703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy     514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
Db     763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy     574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
Db     820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy     631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
Db     880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy     691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
Db     940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy     751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
Db    1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 14
US-07-914-281-1
; Sequence 1, Application US/07914281
; Patent No. 5324663
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/914,281
;      FILING DATE:  19920720
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-07-914-281-1

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 3e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |    | || |     | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
          |   ||  |    | |    |          | |  |  |  |   |||
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453
          |  |||||  |||||  |||   |||||||| || ||    |||||  ||    || |||
Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513
          |    ||          |  |  |  |    |||||  |   |                |
Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573
          || |  |  |   ||||    ||  | ||    |  || || || ||  |||| || ||
Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630
          ||  | |||   || ||||||| || ||||| |     || ||  |   |||  |  |
Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690
          || || || ||  |||||  ||    ||    |||||  ||||  |   | |||  | ||  |
Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750
          |||||  ||  |||||  || |    |    ||||  || || ||   |||  |||  ||||  |
Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795
          |||||  ||  |       |||||  ||  ||||||  ||||  ||  ||
Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

```
RESULT 15
US-08-393-246-1
; Sequence 1, Application US/08393246
; Patent No. 5595900
;  GENERAL INFORMATION:
;    APPLICANT:  LOWE, JOHN B.
;    TITLE OF INVENTION:  METHODS AND PRODUCTS FOR THE SYNTHESIS
;    TITLE OF INVENTION:  OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;    TITLE OF INVENTION:  GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;    TITLE OF INVENTION:  OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,246
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,433
;      FILING DATE:  30-MAR-1994
;      APPLICATION NUMBER:  US 07/914,281
;      FILING DATE:  20-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lavalleye, Jean-Paul M. P.
;      REGISTRATION NUMBER:  31,451
;      REFERENCE/DOCKET NUMBER:  2363-060-55
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-393-246-1

  Query Match             12.3%;  Score 112.2;  DB 1;  Length 2043;
  Best Local Similarity   51.6%;  Pred. No. 3e-24;
  Matches  364;  Conservative    0;  Mismatches  323;  Indels   18;  Gaps    4;

Qy    106 TGCCATCTCACAACGGACCGTTCACTGTACAACAAATCCCATGCAGTTCTGATCCATCAC 165
          |||||  ||||  | |||||     |||||     |  |  |  | ||  |  | || |||
Db    343 TGCCACATCACTGCCGACCGCAAGGTGTACCCACAGGCAGACACGGTCATCGTGCACCAC 402

Qy    166 CGAGACATCAGTTGGGA---TCTGACAAATTTACCTCAGCAAGCTAGGCCACCCTTCCAG 222
          | || ||||  |   |    | || |     | || |       | |||||  |     |||
Db    403 TGGGATATCATGTCCAACCCTAAGTCACGCCTCCCACCTTCCCCGAGGCCGCAGGGGCAG 462

Qy    223 AAATGGATTTGGATGAATTTGGAATCACCAACTCACACTCCCCAAAAGAGTGGCATTGAG 282
            ||||| ||| | || |||||  ||||  || ||   |  ||   |   | | | ||
Db    463 CGCTGGATCTGGTTCAACTTGGAGCCACCCCCTAACTGCCAGCACCTGGAAGCCCTGGAC 522

Qy    283 CACTTGTTTAACCTGACTCTGACTTACCGCCGTGATTCAGATATCCAAGTGCCTTATGGC 342
            |  || || || ||  || | |||||| | || || || |||     ||| || |||
Db    523 AGATACTTCAATCTCACCATGTCCTACCGCAGCGACTCCGACATCTTCACGCCCTACGGC 582

Qy    343 TTCTTGACGGTAAGCACAAATCCCTTCGT---------GTTTGAAGTGCCAAGCAAAGAG 393
```

```
Db    583 TGGCTGGAGCCGTGGTCCGGCCAGCCTGCCCACCCACCGCTCAACCTCTCGGCCAAGACC 642

Qy    394 AAATTGGTGTGCTGGGTTGTGAGTAACTGGAACCCTGAGCATGCCAGAGTCAAGTATTAC 453

Db    643 GAGCTGGTGGCCTGGGCGGTGTCCAACTGGAAGCCGGACTCAGCCAGGGTGCGCTACTAC 702

Qy    454 AATGAGCTAAGCAAAAGCATTGAAATCCATACCTACGGGCAAGCATTTGGAGAATATGTC 513

Db    703 CAGAGCCTGCAGGCTCATCTCAAGGTGGACGTGTACGGACGCTCCCACAAGCCCCTGCCC 762

Qy    514 AATGATAAAAATTTGATTCCTACCATATCTGCTTGTAAATTTTATCTTTCCTTTGAAAAT 573

Db    763 AAGGGGACCA---TGATGGAGACGCTGTCCCGGTACAAGTTCTACCTGGCCTTCGAGAAC 819

Qy    574 TCAATCCACAAGGATTACATCACGGAAAAGCTAT---ACAATGCTTTTCTGGCTGGCTCT 630

Db    820 TCCTTGCACCCCGACTACATCACCGAGAAGCTGTGGAGGAACGCCCTGGAGGCCTGGGCC 879

Qy    631 GTACCTGTTGTTCTGGGACCATCTAGGGAAAACTATGAGAATTATATTCCAGCAGATTCA 690

Db    880 GTGCCCGTGGTGCTGGGCCCCAGCAGAAGCAACTACGAGAGGTTCCTGCCACCCGACGCC 939

Qy    691 TTCATTCATGTGGAAGATTATAACTCTCCCAGTGAGCTAGCAAAGTATCTGAAGGAAGTC 750

Db    940 TTCATCCACGTGGACGACTTCCAGAGCCCCAAGGACCTGGCCCGGTACCTGCAGGAGCTG 999

Qy    751 GACAAAAACAATAAGTTATACCTTAGTTACTTTAACTGGAGGAAG 795

Db   1000 GACAAGGACCACGCCCGCTACCTGAGCTACTTTCGCTGGCGGGAG 1044
```

Search completed: October 6, 2002, 07:22:03
Job time : 30.6694 secs